

From: Swope, Sheridan
Sent: Thursday, October 27, 2005 3:29 PM
To: STIC-Biotech/ChemLib
Subject: 10/649,273

For 10/649,273, please search and interference search:

SID 2: regular search against the NT and AA data bases

SID 2: oligo search (274 residues) against the NT and AA data bases

SID 2: residues 148-414, regular search against the NT and AA data bases

SID 2: residues 176-414, regular search against the NT and AA data bases

Thanks!!

Sheridan Swope, Ph.D.
Patent Examiner, AU 1656
Recombinant Enzymes
571-272-0943 (voice)
E02B71 Remsen Bld (Office)
E03C70 Remsen Bld (Mailbox)

RECEIVED
OCT 27 2005
STIC-BIOTECH/CHEN LIB
(STIC)

Searcher: _____
Searcher Phone: _____
Date Searcher Picked up: _____
Date completed: _____
Searcher Prep Time: _____
Online Time: _____

Type of Search
NA# _____ AA# _____
S/L: _____ Oligomer: _____
Encode/Transl: _____
Structure #: _____ Text: _____
Inventor: _____ Litigation: _____

Vendors and cost where applicable
STN: _____
DIALOG: _____
QUESTEL/ORBIT: _____
LEXIS/NEXIS: _____
SEQUENCE SYSTEM: _____
WWW/Internet: _____
Other (Specify): _____

Swope, Sheridan

From: Swope, Sheridan
Sent: Wednesday, November 02, 2005 9:33 AM
To: Arnold, Deirdre
Subject: RE: 10/649,273

A, match of any contiguous 274 residues of sid 2

-----Original Message-----

From: Arnold, Deirdre
Sent: Wednesday, November 02, 2005 9:10 AM
To: Swope, Sheridan
Subject: RE: 10/649,273

SPL FROM

Do you mean that you are looking for (A) sequences that match 274 contiguous residues of SID 2 or (B) sequences that are at least 274 residues in length and match *any size* portion of SID 2?

Deirdre I. Arnold

United States Patent and Trademark Office
Technical Information Specialist
STIC Biotech/Chem Library
571-272-2532
Remsen 1 A64

-----Original Message-----

From: Swope, Sheridan
Sent: Wednesday, November 02, 2005 9:03 AM
To: Arnold, Deirdre
Subject: FW: 10/649,273

Please communicate via email, so I have a record.

"(274 residues) " means at least 274 residues.
If that's not clear, let me know.

-----Original Message-----

From: Swope, Sheridan
Sent: Thursday, October 27, 2005 3:29 PM
To: STIC-Biotech/ChemLib
Subject: 10/649,273

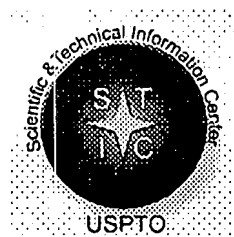
For 10/649,273, please search and interference search:

SID 2: regular search against the NT and AA data bases

SID 2: oligo search (274 residues) against the NT and AA data bases

SID 2: residues 148-414, regular search against the NT and AA data bases

SID 2: residues 176-414, regular search against the NT and AA data bases



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 169778

TO: Sheridan Swope
Location: REM/2B71/3C70
Art Unit: 1656
Monday, November 14, 2005

Case Serial Number: 10/649273

From: Deirdre Arnold
Location: Biotech-Chem Library
REM 1A64
Phone: 571-272-2532

Deirdre.Arnold@uspto.gov

Search Notes

Please review these results and contact me ASAP if you need to alter the search parameters.

Please feel free to contact me if you have any questions or would like to amend the search.

Thank you for using STIC services.

Regards,
Deirdre Arnold



GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2005, 16:43:58 ; Search time 109.35 Seconds

(without alignment)
1584.102 Million cell updates/sec

Title: US-10-649-273-2

Perfect score: 1 MLILTKTAGVFPKSKRKRY.....DISKEVGASIKVQLKMEI 414

Sequence: 1 MLILTKTAGVFPKSKRKRY.....DISKEVGASIKVQLKMEI 414

Scoring table: BLOSUM62

Searched: Gap0 10.0 , Gapext 0.5

Total number of hits satisfying chosen parameters: 1867879

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*

2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*

3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*

4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*

5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*

6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*

7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*

8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*

9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*

10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*

11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*

12: /cgn2_6/ptodata/1/pubpaa/US09C_NEW_PUB.pep.*

13: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*

14: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*

15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*

16: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*

17: /cgn2_6/ptodata/1/pubpaa/US10E_PUBCOMB.pep.*

18: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*

19: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep.*

20: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*

21: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*

22: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2125	100.0	414	US-10-067-443-2	Sequence 2, Appli
2	2125	100.0	414	US-10-649-273-2	Sequence 2, Appli
3	2125	100.0	414	US-10-651-722-2	Sequence 2, Appli
4	2125	100.0	414	US-10-480-988-8	Sequence 8, Appli
5	2090.5	98.4	439	US-10-067-443-19	Sequence 19, Appli
6	2090.5	98.4	439	US-10-649-273-19	Sequence 19, Appli
7	2090.5	98.4	439	US-10-651-722-19	Sequence 19, Appli
8	2088	98.3	414	US-10-012-140-5	Sequence 5, Appli
9	1845	86.8	364	US-10-094-749-2039	Sequence 2039, Ap
10	1385	65.2	267	US-10-067-443-22	Sequence 22, Appli
11	1385	65.2	267	US-10-649-273-22	Sequence 22, Appli

12	1385	65.2	267	15	US-10-651-722-22	Sequence 22, Appli
13	714.5	33.6	409	20	US-11-097-143-34191	Sequence 34191, A
14	681.5	32.1	445	15	US-10-424-599-209259	Sequence 209259, A
15	634	29.8	463	14	US-10-067-443-3	Sequence 3, Appli
16	634	29.8	463	15	US-10-649-273-3	Sequence 3, Appli
17	634	29.8	463	15	US-10-651-722-3	Sequence 3, Appli
18	578	27.2	444	16	US-10-437-963-113732	Sequence 113732, A
19	556.5	26.2	333	14	US-10-012-140-25	Sequence 25, Appli
20	549	25.8	179	14	US-10-067-443-25	Sequence 25, Appli
21	549	25.8	179	15	US-10-649-273-25	Sequence 25, Appli
22	549	25.8	179	15	US-10-651-722-25	Sequence 25, Appli
23	524	24.7	382	15	US-10-282-122A-50858	Sequence 50858, A
24	524	24.7	421	14	US-10-067-443-4	Sequence 4, Appli
25	524	24.7	421	14	US-10-067-443-28	Sequence 28, Appli
26	524	24.7	421	15	US-10-649-273-4	Sequence 4, Appli
27	524	24.7	421	15	US-10-649-273-28	Sequence 28, Appli
28	524	24.7	421	15	US-10-651-722-4	Sequence 4, Appli
29	524	24.7	421	15	US-10-651-722-28	Sequence 28, Appli
30	511.5	24.1	347	14	US-10-012-140-24	Sequence 24, Appli
31	502	23.6	348	15	US-10-282-122A-63156	Sequence 63156, A
32	494.5	23.3	343	15	US-10-282-122A-67227	Sequence 67227, A
33	492.5	23.2	341	9	US-09-815-242-11798	Sequence 11798, A
34	492.5	23.2	341	15	US-10-282-122A-66200	Sequence 66200, A
35	492.5	23.2	341	18	US-10-958-216-50	Sequence 50, Appli
36	490.5	23.1	341	18	US-10-958-216-52	Sequence 52, Appli
37	489.5	23.0	361	15	US-10-282-122A-55404	Sequence 55404, A
38	488.5	23.0	361	16	US-10-425-115-352095	Sequence 352095, A
39	484.5	22.8	337	9	US-09-815-242-10304	Sequence 10304, A
40	484.5	22.8	337	15	US-10-282-122A-56695	Sequence 56695, A
41	484.5	22.8	337	15	US-10-282-122A-75485	Sequence 75485, A
42	483.5	22.8	337	15	US-10-282-122A-78161	Sequence 78161, A
43	482.5	22.7	340	15	US-10-282-122A-68438	Sequence 68438, A
44	481.5	22.7	337	9	US-09-815-242-113780	Sequence 113780, A
45	475.5	22.4	342	9	US-09-815-242-11043	Sequence 11043, A

ALIGNMENTS

RESULT 1

US-10-067-443-2

Sequence 2, Application US/10067443

Publication No. US20030082782A1

GENERAL INFORMATION:

APPLICANT: Bristol-Myers Squibb Company

TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN SPINAL CORD, MP-1

FILE REFERENCE: D0073 NP

CURRENT APPLICATION NUMBER: US/10/067,443

CURRENT FILING DATE: 2002-02-05

PRIOR APPLICATION NUMBER: US 60/266,518

PRIOR FILING DATE: 2001-02-05

PRIOR APPLICATION NUMBER: US 60/282,814

PRIOR FILING DATE: 2001-04-10

NUMBER OF SEQ ID NOS: 71

SOFTWARE: PatentIn version 3.0

SEQ ID NO 2

LENGTH: 414

TYPE: PRT

ORGANISM: Homo sapiens

US-10-067-443-2

Query Match 100.0%; Score 2125; DB 14; Length 414;
Best Local Similarity 100.0%; Pred. No. 1.7e-199;
Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLILTKTAGVFPKSKRKRYEFLRSPNFRGTLFLKIVLGIETSCDDTAAVDETGV 60

DB 1 MLILTKTAGVFPKSKRKRYEFLRSPNFRGTLFLKIVLGIETSCDDTAAVDETGV 60

QY 61 LGEALHSQTEVTLKNGVILPPAQQLRNRIYQVQALSNAGVSPSDLSAATIKPL 120

DB 61 LGEALHSQTEVTLKNGVILPPAQQLRNRIYQVQALSNAGVSPSDLSAATIKPL 120

QY 121 ALSIGVGLSFSLOLVGKPKPFIPIHMEAHALITRLTNKVEPFLVLLISGHCILALV 180
DB 121 ALSIGVGLSFSLOLVGKPKPFIPIHMEAHALITRLTNKVEPFLVLLISGHCILALV 180
QY 181 QGVSDPFLILGKSLDIAPGDMLDKVARRLSLIKHPECSTMSGKAIIEHLAKQGNRPHFDIK 240
DB 181 QGVSDPFLILGKSLDIAPGDMLDKVARRLSLIKHPECSTMSGKAIIEHLAKQGNRPHFDIK 240
QY 241 PPLHAKKCDPSFTGLQHVTDKTIIMKKEKEBEGIEKGQILSSAADIAATVOHTMACHLVYK 300
DB 241 PPLHAKKCDPSFTGLQHVTDKTIIMKKEKEBEGIEKGQILSSAADIAATVOHTMACHLVYK 300
QY 301 THRAILFCQKRDLLPQNNNAVLVASGVASNFYIRALBELITNAOCTLLCPPRLCTDNG 360
DB 301 THRAILFCQKRDLLPQNNNAVLVASGVASNFYIRALBELITNAOCTLLCPPRLCTDNG 360
QY 361 IMIANGIERLRAGLGIHIDIEGIRYBPKCPGLVDISKVEGASIKVPOLKMEI 414
DB 361 IMIANGIERLRAGLGIHIDIEGIRYBPKCPGLVDISKVEGASIKVPOLKMEI 414

RESULT 2
US-10-649-273-2
; Sequence 2, Application US/10649273
; Publication No. US20040043407A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1
; FILE REFERENCE: D0073 CNT
; CURRENT APPLICATION NUMBER: US/10/649,273
; PRIOR FILING DATE: 2003-08-27
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-649-273-2

Query Match 100.0%; Score 2125; DB 15; Length 414;
Best Local Similarity 100.0%; Pred. No. 1.7e-199;
Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MLITKTAGVFFPKSRKRYEFLRSFNPHGTLFLHKIVLGITSCDDTAAAVDETGNV 60
DB 1 MLITKTAGVFFPKSRKRYEFLRSFNPHGTLFLHKIVLGITSCDDTAAAVDETGNV 60
QY 61 LGEAHSOTEVHLKTTGGIVPPAAQOLHRENIQRIVOEALSASGVSPDSLSAATTKIRGL 120
DB 61 LGEAHSOTEVHLKTTGGIVPPAAQOLHRENIQRIVOEALSASGVSPDSLSAATTKIRGL 120
QY 121 ALSIGVGLSFSLOLVGKPKPFIPIHMEAHALITRLTNKVEPFLVLLISGHCILALV 180
DB 121 ALSIGVGLSFSLOLVGKPKPFIPIHMEAHALITRLTNKVEPFLVLLISGHCILALV 180
QY 181 QGVSDPFLILGKSLDIAPGDMLDKVARRLSLIKHPECSTMSGKAIIEHLAKQGNRPHFDIK 240
DB 181 QGVSDPFLILGKSLDIAPGDMLDKVARRLSLIKHPECSTMSGKAIIEHLAKQGNRPHFDIK 240
QY 241 PPLHAKKCDPSFTGLQHVTDKTIIMKKEKEBEGIEKGQILSSAADIAATVOHTMACHLVYK 300
DB 241 PPLHAKKCDPSFTGLQHVTDKTIIMKKEKEBEGIEKGQILSSAADIAATVOHTMACHLVYK 300
QY 301 THRAILFCQKRDLLPQNNNAVLVASGVASNFYIRALBELITNAOCTLLCPPRLCTDNG 360
DB 301 THRAILFCQKRDLLPQNNNAVLVASGVASNFYIRALBELITNAOCTLLCPPRLCTDNG 360

QY 361 IMIANGIERLRAGLGIHIDIEGIRYBPKCPGLVDISKVEGASIKVPOLKMEI 414
DB 361 IMIANGIERLRAGLGIHIDIEGIRYBPKCPGLVDISKVEGASIKVPOLKMEI 414

RESULT 3
US-10-651-722-2
; Sequence 2, Application US/10651722
; Publication No. US20040048302A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1
; FILE REFERENCE: D0073 DIV
; CURRENT APPLICATION NUMBER: US/10/651,722
; PRIOR FILING DATE: 2003-08-29
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-651-722-2

Query Match 100.0%; Score 2125; DB 15; Length 414;
Best Local Similarity 100.0%; Pred. No. 1.7e-199;
Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLITKTAGVFFPKSRKRYEFLRSFNPHGTLFLHKIVLGITSCDDTAAAVDETGNV 60
DB 1 MLITKTAGVFFPKSRKRYEFLRSFNPHGTLFLHKIVLGITSCDDTAAAVDETGNV 60
QY 61 LGEAHSOTEVHLKTTGGIVPPAAQOLHRENIQRIVOEALSASGVSPDSLSAATTKIRGL 120
DB 61 LGEAHSOTEVHLKTTGGIVPPAAQOLHRENIQRIVOEALSASGVSPDSLSAATTKIRGL 120
QY 121 ALSIGVGLSFSLOLVGKPKPFIPIHMEAHALITRLTNKVEPFLVLLISGHCILALV 180
DB 121 ALSIGVGLSFSLOLVGKPKPFIPIHMEAHALITRLTNKVEPFLVLLISGHCILALV 180
QY 181 QGVSDPFLILGKSLDIAPGDMLDKVARRLSLIKHPECSTMSGKAIIEHLAKQGNRPHFDIK 240
DB 181 QGVSDPFLILGKSLDIAPGDMLDKVARRLSLIKHPECSTMSGKAIIEHLAKQGNRPHFDIK 240
QY 241 PPLHAKKCDPSFTGLQHVTDKTIIMKKEKEBEGIEKGQILSSAADIAATVOHTMACHLVYK 300
DB 241 PPLHAKKCDPSFTGLQHVTDKTIIMKKEKEBEGIEKGQILSSAADIAATVOHTMACHLVYK 300
QY 301 THRAILFCQKRDLLPQNNNAVLVASGVASNFYIRALBELITNAOCTLLCPPRLCTDNG 360
DB 301 THRAILFCQKRDLLPQNNNAVLVASGVASNFYIRALBELITNAOCTLLCPPRLCTDNG 360
QY 361 IMIANGIERLRAGLGIHIDIEGIRYBPKCPGLVDISKVEGASIKVPOLKMEI 414
DB 361 IMIANGIERLRAGLGIHIDIEGIRYBPKCPGLVDISKVEGASIKVPOLKMEI 414

RESULT 4
US-10-480-988-8
; Sequence 8, Application US/10480988
; Publication No. US20050069877A1
; GENERAL INFORMATION:
; APPLICANT: GANDHI, Ameena R.; KABLE, Amy E.;
; APPLICANT: SWARNAKAR, Anita; HAFALITA, April J.A.;
; APPLICANT: TRAN, Bao; DUGGAN, Brendan M.;
; APPLICANT: WARREN, Bridget A.; ISON, Craig H.;
; APPLICANT: HONCHELL, Cynthia D.; NGUYEN, Daniel B.;

```
; APPLICANT: LU, Dyrung Aina M.; LEE, Ernestine A.;
; APPLICANT: YUE, Henry; FORSTNER, Ian J.;
; APPLICANT: BARROSO, Ines; RAMKUMAR, Jayalaxmi;
; APPLICANT: GRIFFIN, Jennifer A.; LI, Joana X.;
; APPLICANT: YANG, Junming; THANGAVELU, Kavitha;
; APPLICANT: GIETZEN, Kimberly J.; DING, Li;
; APPLICANT: BAUGHN, Mariah R.; BOROMSKY, Mark L.;
; APPLICANT: YAO, Monique G.; CHAMLA, Narinder K.;
; APPLICANT: MASON, Patricia M.; GURURAJAN, Rajagopal;
; APPLICANT: LEE, Sally; BECHA, Shanya D.;
; APPLICANT: LEE, Soo Yeun; TRAN, Uyen K.;
; APPLICANT: ELLIOTT, Vicki S.; LUO, Wen;
; APPLICANT: SPRAGUE, William W.; TANG, Y. Tom;
; APPLICANT: LU, Yan; ZEBARJADIAN, Yeganeh
; TITLE OF INVENTION: PROTEIN MODIFICATION AND MAINTENANCE MOLECULES
; FILE REFERENCE: PF-1040 USN
; CURRENT APPLICATION NUMBER: US/10/480,988
; PRIOR FILING DATE: 2003-12-16
; PRIOR APPLICATION NUMBER: PCT/US02/19360
; PRIOR FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: US 60/300,508
; PRIOR FILING DATE: 2001-06-22
; PRIOR APPLICATION NUMBER: US 60/303,445
; PRIOR FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US 60/305,405
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/311,442
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: US 60/314,821
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/315,992
; PRIOR FILING DATE: 2001-08-29
; PRIOR APPLICATION NUMBER: US 60/378,205
; PRIOR FILING DATE: 2002-05-03
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: PERL Program
; SEQ ID NO 8
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: 7632424CD1
US-10-480-988-8

Query Match      100.0%; Score 2125; DB 17; Length 414;
Best Local Similarity 100.0%; Pred. No. 1,7e-199;
Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 MLILTKAGVFPKSKRYEFLRSFNFPGTLFLHKIVLGIRTSCTDPTAAAVDENGTV 60
DB 1 MLILTKAGVFPKSKRYEFLRSFNFPGTLFLHKIVLGIRTSCTDPTAAAVDENGTV 60
QY 61 LGEAHSQTEVHLKTGGIVPPAAOQLHRENTORIVQEAALSASGVSPDLSAATTIKPGL 120
DB 61 LGEAHSQTEVHLKTGGIVPPAAOQLHRENTORIVQEAALSASGVSPDLSAATTIKPGL 120
QY 121 ALSIGVLSFSIQLVGOLKKPPIPIHMEAHALTIRLTNKEPPLVLLISGHCILALV 180
DB 121 ALSIGVLSFSIQLVGOLKKPPIPIHMEAHALTIRLTNKEPPLVLLISGHCILALV 180
QY 181 QGVSDPFLILGKSLDIAPGMDLVKVARRLSLIKHPECSTWSGGKAIETHLAKQGNRFHFDIK 240
DB 181 QGVSDPFLILGKSLDIAPGMDLVKVARRLSLIKHPECSTWSGGKAIETHLAKQGNRFHFDIK 240
QY 241 PPLHAKNCDPSFTGLQHTVDKIIMKKEKEGIEKGQILSSAADIATVQHTMACHLVYR 300
DB 241 PPLHAKNCDPSFTGLQHTVDKIIMKKEKEGIEKGQILSSAADIATVQHTMACHLVYR 300
QY 301 TPEALIFCCKQSDLLPQNNAVLVASGVASNFYIRSLALITNAVCTLLCPEPKLCTUNG 360
DB 301 TPEALIFCCKQSDLLPQNNAVLVASGVASNFYIRSLALITNAVCTLLCPEPKLCTUNG 360
```

```
QY 361 IMIANGIRLRLAGLILHIDIEGIRYEPKCPGLVDISKEVGASIKVPOLKMEI 414
DB 361 IMIANGIRLRLAGLILHIDIEGIRYEPKCPGLVDISKEVGASIKVPOLKMEI 414

RESULT 5
US-10-067-443-19
; Sequence 19, Application US/10067443
; Publication No. US20030082782A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 19
; LENGTH: 439
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-067-443-19

Query Match      98.4%; Score 2090.5; DB 14; Length 439;
Best Local Similarity 93.8%; Pred. No. 4.6e-196;
Matches 412; Conservative 0; Mismatches 2; Indels 25; Gaps 1;
```

```
QY 1 MLILTKAGVFPKSKRYEFLRSFNFPGTLFLHKIVLGIRTSCTDPTAAAVDENGTV 60
DB 1 MLILTKAGVFPKSKRYEFLRSFNFPGTLFLHKIVLGIRTSCTDPTAAAVDENGTV 60
QY 61 LGEAHSQTEVHLKTGGIVPPAAOQLHRENTORIVQEAALSASGVSPDLSAATTIKPGL 120
DB 61 LGEAHSQTEVHLKTGGIVPPAAOQLHRENTORIVQEAALSASGVSPDLSAATTIKPGL 120
QY 121 ALSIGVLSFSIQLVGOLKKPPIPIHMEAHALTIRLTNKEPPLVLLISGHCILALV 180
DB 121 ALSIGVLSFSIQLVGOLKKPPIPIHMEAHALTIRLTNKEPPLVLLISGHCILALV 180
QY 181 QGVSDPFLILGKSLDIAPGMDLVKVARRLSLIKHPECSTWSGGKAIETHLAKQGNRFHFDIK 240
DB 181 QGVSDPFLILGKSLDIAPGMDLVKVARRLSLIKHPECSTWSGGKAIETHLAKQGNRFHFDIK 240
QY 241 PPLHAKNCDPSFTGLQHTVDKIIMKKEKEGIEKGQILSSAADIATVQHTMACHLVYR 300
DB 241 PPLHAKNCDPSFTGLQHTVDKIIMKKEKEGIEKGQILSSAADIATVQHTMACHLVYR 300
QY 301 TPEALIFCCKQSDLLPQNNAVLVASGVASNFYIRSLALITNAVCTLLCPEPKLCTUNG 360
DB 301 TPEALIFCCKQSDLLPQNNAVLVASGVASNFYIRSLALITNAVCTLLCPEPKLCTUNG 360
QY 361 IMIANGIRLRLAGLILHIDIEGIRYEPKCPGLVDISKEVGASIKVPOLKMEI 414
DB 361 IMIANGIRLRLAGLILHIDIEGIRYEPKCPGLVDISKEVGASIKVPOLKMEI 414

RESULT 6
US-10-649-273-19
; Sequence 19, Application US/10649273
; Publication No. US20040043407A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1
; FILE REFERENCE: D0073 CNT
```

```
; CURRENT APPLICATION NUMBER: US/10/649, 273
; CURRENT FILING DATE: 2003-08-27
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 19
; LENGTH: 439
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-649-273-19

Query Match      98.4%; Score 2090.5; DB 15; Length 439;
Best Local Similarity 93.8%; Pred. No. 4.6e-196;
Matches 412; Conservative 0; Mismatches 2; Indels 25; Gaps 1;

QY 1 MLILTKTAGVFFPKSKRYEFLRSFNFPGTLFLHKIVLGIESCDPTAAAVDETVGNV 60
DB 1 MLILTKTAGVFFPKSKRYEFLRSFNFPGTLFLHKIVLGIESCDPTAAAVDETVGNV 60
QY 61 LGEAHSQTEVHLKTGGIVPPAAQOLHRENIQRIVOEALASAGVSPSDLSAIAITTIKPGI 120
DB 61 LGEAHSQTEVHLKTGGIVPPAAQOLHRENIQRIVOEALASAGVSPSDLSAIAITTIKPGI 120
QY 121 ALISGVLSFSIQVLVQGLKKPPIPIHMEAHALITRLTNKVEPPLVLLISGHCILLALV 180
DB 121 ALISGVLSFSIQVLVQGLKKPPIPIHMEAHALITRLTNKVEPPLVLLISGHCILLALV 180
QY 181 QGVDFLLGKSLDIPGMDLKVARRSLIKHPECSTMSGKAI EHLAKQNRFFFDIK 240
DB 181 QGVDFLLGKSLDIPGMDLKVARRSLIKHPECSTMSGKAI EHLAKQNRFFFDIK 240
QY 241 PPLHAKKNCDFSTGLQHTVTDKIIMKKEKEGII-----EK 275
DB 241 PPLHAKKNCDFSTGLQHTVTDKIIMKKEKEGII-----EK 275
QY 276 GQILSSADIAATVQHTMACHLVKTRHAILFCQKRDLLPQNNAVLVASGVASNFYIRR 335
DB 301 GQILSSADIAATVQHTMACHLVKTRHAILFCQKRDLLPQNNAVLVASGVASNFYIRR 360
QY 336 ALBITLNATQCTLLCPPEPLCTDNGIMIAMNGIERLRAGLIGLHDI EGRYEPKCPGLVD 395
DB 361 ALBITLNATQCTLLCPPEPLCTDNGIMIAMNGIERLRAGLIGLHDI EGRYEPKCPGLVD 420
QY 396 ISKEVGEASIKVPOLKMEI 414
DB 421 ISKEVGEASIKVPOLKMEI 439

RESULT 7
US-10-651-722-19
; Sequence 19, Application US/10651722
; Publication No. US20040048302A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1
; FILE REFERENCE: D0073 DIV
; CURRENT APPLICATION NUMBER: US/10/651,722
; PRIOR FILING DATE: 2003-08-29
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 19
; LENGTH: 439
```

```
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-651-722-19

Query Match      98.4%; Score 2090.5; DB 15; Length 439;
Best Local Similarity 93.8%; Pred. No. 4.6e-196;
Matches 412; Conservative 0; Mismatches 2; Indels 25; Gaps 1;

QY 1 MLILTKTAGVFFPKSKRYEFLRSFNFPGTLFLHKIVLGIESCDPTAAAVDETVGNV 60
DB 1 MLILTKTAGVFFPKSKRYEFLRSFNFPGTLFLHKIVLGIESCDPTAAAVDETVGNV 60
QY 61 LGEAHSQTEVHLKTGGIVPPAAQOLHRENIQRIVOEALASAGVSPSDLSAIAITTIKPGI 120
DB 61 LGEAHSQTEVHLKTGGIVPPAAQOLHRENIQRIVOEALASAGVSPSDLSAIAITTIKPGI 120
QY 121 ALISGVLSFSIQVLVQGLKKPPIPIHMEAHALITRLTNKVEPPLVLLISGHCILLALV 180
DB 121 ALISGVLSFSIQVLVQGLKKPPIPIHMEAHALITRLTNKVEPPLVLLISGHCILLALV 180
QY 181 QGVDFLLGKSLDIPGMDLKVARRSLIKHPECSTMSGKAI EHLAKQNRFFFDIK 240
DB 181 QGVDFLLGKSLDIPGMDLKVARRSLIKHPECSTMSGKAI EHLAKQNRFFFDIK 240
QY 241 PPLHAKKNCDFSTGLQHTVTDKIIMKKEKEGII-----EK 275
DB 241 PPLHAKKNCDFSTGLQHTVTDKIIMKKEKEGII-----EK 275
QY 276 GQILSSADIAATVQHTMACHLVKTRHAILFCQKRDLLPQNNAVLVASGVASNFYIRR 335
DB 301 GQILSSADIAATVQHTMACHLVKTRHAILFCQKRDLLPQNNAVLVASGVASNFYIRR 360
QY 336 ALBITLNATQCTLLCPPEPLCTDNGIMIAMNGIERLRAGLIGLHDI EGRYEPKCPGLVD 395
DB 361 ALBITLNATQCTLLCPPEPLCTDNGIMIAMNGIERLRAGLIGLHDI EGRYEPKCPGLVD 420
QY 396 ISKEVGEASIKVPOLKMEI 414
DB 421 ISKEVGEASIKVPOLKMEI 439

RESULT 8
US-10-012-140-5
; Sequence 5, Application US/10012140
; Publication No. US20030009017A1
; GENERAL INFORMATION:
; APPLICANT: Leiby, Kevin R.
; APPLICANT: Kappeller-Libermann, Rosana
; TITLE OF INVENTION: 38650, 28472, 5495, 65507, 81588, AND
; TITLE OF INVENTION: 14354 METHODS AND COMPOSITIONS OF HUMAN PROTEINS AND USES
; FILE REFERENCE: 381552004900
; CURRENT APPLICATION NUMBER: US/10/012,140
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: 60/246,768
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/246,772-yes
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/249,185
; PRIOR FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-012-140-5

Query Match      98.3%; Score 2088; DB 14; Length 414;
Best Local Similarity 98.3%; Pred. No. 7.5e-196;
Matches 407; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
```

QY 1 MLITKTAGVPEKPSKRVYEFLEBSFNPHGTLFLHKTIVLGLETSCDDTAAAVDDEGNV 60
DB 1 MLITKTAGVPEKPSKRVYEFLEBSFNPHGTLFLHKTIVLGLETSCDDTAAAVDDEGNV 60
QY 61 LGEAHSQTEVHLKTGIVPPAAQOLHRENIOIYQVQALASASVSPSDLSAATTIKPGL 120
DB 61 LGEAHSQTEVHLKTGIVPPAAQOLHRENIOIYQVQALASASVSPSDLSAATTIKPGL 120
QY 121 ALSIGVGSFSLQVLQVQKKPPIPIHMEAHALTRLTNKVEPPLVLLISGCHLLAV 180
DB 121 ALSIGVGSFSLQVLQVQKKPPIPIHMEAHALTRLTNKVEPPLVLLISGCHLLAV 180
QY 181 QGVSDPFLLGKSLDIPAGMDLVKVARRLSLIKPECSSTMSGGKAIHHLAKQGNRPFHDIX 240
DB 181 QGVSDPFLLGKSLDIPAGMDLVKVARRLSLIKPECSSTMSGGKAIHHLAKQGNRPFHDIX 240
QY 241 PPLHAKNCDPFTGLQHTVTDKIIMKKEBEGIEKQIILSSADIAATVOHTMACHLVVR 300
DB 241 PPLHAKNCDPFTGLQHTVTDKIIMKKEBEGIEKQIILSSADIAATVOHTMACHLVVR 300
QY 301 THRAILFCQRDLPPONNAVLVASGVASNFYIRRALBITNATQCTLLCPPRLCTDNG 360
DB 301 THRAILFCQRDLPPONNAVLVASGVASNFYIRRALBITNATQCTLLCPPRLCTDNG 360
QY 361 IMIANGIERLAGIGIHDIEGIRYEPKCPGVDISKEVGEASIKVPOLKMEI 414
DB 361 IMIANGIERLAGIGIHDIEGIRYEPKCPGVDISKEVGEASIKVPOLKMEI 414

RESULT 9
US-10-094-749-2039
; Sequence 2039, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOKYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; PRIOR FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO: 2039
; LENGTH: 364
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-749-2039

Query Match 86.8%; Score 1845; DB 15; Length 364;
Best Local Similarity 93.2%; Pred. No. 4, 5e-172;
Matches 361; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 MLITKTAGVPEKPSKRVYEFLEBSFNPHGTLFLHKTIVLGLETSCDDTAAAVDDEGNV 60

DB 1 MLITKTAGVPEKPSKRVYEFLEBSFNPHGTLFLHKTIVLGLETSCDDTAAAVDDEGNV 60
QY 61 LGEAHSQTEVHLKTGIVPPAAQOLHRENIOIYQVQALASASVSPSDLSAATTIKPGL 120
DB 61 LGEAHSQTEVHLKTGIVPPAAQOLHRENIOIYQVQALASASVSPSDLSAATTIKPGL 120
QY 121 ALSIGVGSFSLQVLQVQKKPPIPIHMEAHALTRLTNKVEPPLVLLISGCHLLAV 180
DB 121 ALSIGVGSFSLQVLQVQKKPPIPIHMEAHALTRLTNKVEPPLVLLISGCHLLAV 180
QY 181 QGVSDPFLLGKSLDIPAGMDLVKVARRLSLIKPECSSTMSGGKAIHHLAKQGNRPFHDIX 240
DB 181 QGVSDPFLLGKSLDIPAGMDLVKVARRLSLIKPECSSTMSGGKAIHHLAKQGNRPFHDIX 240
QY 241 PPLHAKNCDPFTGLQHTVTDKIIMKKEBEGIEKQIILSSADIAATVOHTMACHLVVR 300
DB 241 PPLHAKNCDPFTGLQHTVTDKIIMKKEBEGIEKQIILSSADIAATVOHTMACHLVVR 300
QY 301 THRAILFCQRDLPPONNAVLVASGVASNFYIRRALBITNATQCTLLCPPRLCTDNG 360
DB 301 THRAILFCQRDLPPONNAVLVASGVASNFYIRRALBITNATQCTLLCPPRLCTDNG 360
QY 361 IMIA 364
DB 361 IMIA 364

RESULT 10
US-10-067-443-22
; Sequence 22, Application US/10067443
; Publication No. US20030082782A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED I
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO: 22
; LENGTH: 267
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-067-443-22

Query Match 65.2%; Score 1385; DB 14; Length 267;
Best Local Similarity 100.0%; Pred. No. 4, 2e-127;
Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 148 MEAHALTRLTNKVEPPLVLLISGCHLLAVQGVSDPFLLGKSLDIPAGMDLVKVAR 207
DB 1 MEAHALTRLTNKVEPPLVLLISGCHLLAVQGVSDPFLLGKSLDIPAGMDLVKVAR 60
QY 208 LSLIKPECSSTMSGGKAIHHLAKQGNRPFHDIXPPLHAKNCDPFTGLQHTVTDKIIMK 267
DB 208 LSLIKPECSSTMSGGKAIHHLAKQGNRPFHDIXPPLHAKNCDPFTGLQHTVTDKIIMK 120
QY 268 EKEBEGIEKQIILSSADIAATVOHTMACHLVKTTHRAILFCQRDLPPONNAVLVASGV 327
DB 268 EKEBEGIEKQIILSSADIAATVOHTMACHLVKTTHRAILFCQRDLPPONNAVLVASGV 180
QY 121 EKEBEGIEKQIILSSADIAATVOHTMACHLVKTTHRAILFCQRDLPPONNAVLVASGV 180
DB 121 EKEBEGIEKQIILSSADIAATVOHTMACHLVKTTHRAILFCQRDLPPONNAVLVASGV 180
QY 328 ASNFYIRRALBITNATQCTLLCPPRLCTDNGIMIANGIERLAGIGIHDIEGIRY 387
DB 328 ASNFYIRRALBITNATQCTLLCPPRLCTDNGIMIANGIERLAGIGIHDIEGIRY 240
QY 368 PKCPLGVDSKEVGEASIKVPOLKMEI 414
DB 368 PKCPLGVDSKEVGEASIKVPOLKMEI 414

Db 241 PKCPLGVDISKVEGASIKVPOLKMEI 267

RESULT 11

US-10-649-273-22

Sequence 22, Application US/10649273

Publication No. US20040043407A1

GENERAL INFORMATION:

APPLICANT: Bristol-Myers Squibb Company

TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1

FILE REFERENCE: D0073 CNT

CURRENT APPLICATION NUMBER: US/10/649,273

PRIOR FILING DATE: 2003-08-27

PRIOR APPLICATION NUMBER: US 60/266,518

PRIOR FILING DATE: 2001-02-05

PRIOR APPLICATION NUMBER: US 10/067,443

PRIOR FILING DATE: 2002-02-05

PRIOR APPLICATION NUMBER: US 60/282,814

PRIOR FILING DATE: 2001-04-10

NUMBER OF SEQ ID NOS: 71

SOFTWARE: PatentIn version 3.2

SEQ ID NO 22

LENGTH: 267

TYPE: PRT

ORGANISM: homo sapiens

US-10-649-273-22

Query Match 65.2%; Score 1385; DB 15; Length 267;

Best Local Similarity 100.0%; Pred. No. 4.2e-127;

Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 148 MEAHALTRLTNKVEFPFLVLLISGGHCLLALVGVSDFLLGKSLDIAFGDMLDKVARR 207

Db 1 MEAHALTRLTNKVEFPFLVLLISGGHCLLALVGVSDFLLGKSLDIAFGDMLDKVARR 60

Qy 208 LSLIKHPECSTMSGKAIIEHLAKQGNRFHFDIKPPLHAKNCDPFTGLOHTYDXTIMKK 267

Db 61 LSLIKHPECSTMSGKAIIEHLAKQGNRFHFDIKPPLHAKNCDPFTGLOHTYDXTIMKK 120

Qy 268 EKEBIEGQILSSADIAATVQHTMACHLVKTRTRAILFCQORDLLPNNNAVVASGGV 327

Db 121 EKEBIEGQILSSADIAATVQHTMACHLVKTRTRAILFCQORDLLPNNNAVVASGGV 180

Qy 328 ASNFYIRRALIELTNATQCTLLCPPRLCTDNGIMIANWGIERLRAGLGIHDIISIRYE 387

Db 181 ASNFYIRRALIELTNATQCTLLCPPRLCTDNGIMIANWGIERLRAGLGIHDIISIRYE 240

Qy 388 PKCPLGVDISKVEGASIKVPOLKMEI 414

Db 241 PKCPLGVDISKVEGASIKVPOLKMEI 267

RESULT 12

US-10-651-722-22

Sequence 22, Application US/10651722

Publication No. US20040048302A1

GENERAL INFORMATION:

APPLICANT: Bristol-Myers Squibb Company

TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1

FILE REFERENCE: D0073 DIV

CURRENT APPLICATION NUMBER: US/10/651,722

PRIOR FILING DATE: 2003-08-29

PRIOR APPLICATION NUMBER: US 60/266,518

PRIOR FILING DATE: 2001-02-05

PRIOR APPLICATION NUMBER: US 10/067,443

PRIOR FILING DATE: 2002-02-05

PRIOR APPLICATION NUMBER: US 60/282,814

PRIOR FILING DATE: 2001-04-10

NUMBER OF SEQ ID NOS: 71

SOFTWARE: PatentIn version 3.2

SEQ ID NO 22

LENGTH: 267

TYPE: PRT

ORGANISM: homo sapiens

US-10-651-722-22

Query Match 65.2%; Score 1385; DB 15; Length 267;

Best Local Similarity 100.0%; Pred. No. 4.2e-127;

Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 148 MEAHALTRLTNKVEFPFLVLLISGGHCLLALVGVSDFLLGKSLDIAFGDMLDKVARR 207

Db 1 MEAHALTRLTNKVEFPFLVLLISGGHCLLALVGVSDFLLGKSLDIAFGDMLDKVARR 60

Qy 208 LSLIKHPECSTMSGKAIIEHLAKQGNRFHFDIKPPLHAKNCDPFTGLOHTYDXTIMKK 267

Db 61 LSLIKHPECSTMSGKAIIEHLAKQGNRFHFDIKPPLHAKNCDPFTGLOHTYDXTIMKK 120

Qy 268 EKEBIEGQILSSADIAATVQHTMACHLVKTRTRAILFCQORDLLPNNNAVVASGGV 327

Db 121 EKEBIEGQILSSADIAATVQHTMACHLVKTRTRAILFCQORDLLPNNNAVVASGGV 180

Qy 328 ASNFYIRRALIELTNATQCTLLCPPRLCTDNGIMIANWGIERLRAGLGIHDIISIRYE 387

Db 181 ASNFYIRRALIELTNATQCTLLCPPRLCTDNGIMIANWGIERLRAGLGIHDIISIRYE 240

Qy 388 PKCPLGVDISKVEGASIKVPOLKMEI 414

Db 241 PKCPLGVDISKVEGASIKVPOLKMEI 267

RESULT 13

US-11-097-143-34191

Sequence 34191, Application US/11097143

Publication No. US20050208558A1

GENERAL INFORMATION:

APPLICANT: Venter, J. Craig

TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID

TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE

FILE REFERENCE: C1000728

CURRENT APPLICATION NUMBER: US/11/097,143

PRIOR FILING DATE: 2005-04-04

PRIOR APPLICATION NUMBER: 60/157,832

PRIOR FILING DATE: 1999-10-05

PRIOR APPLICATION NUMBER: 60/160,191

PRIOR FILING DATE: 1999-10-19

PRIOR APPLICATION NUMBER: 60/161,932

PRIOR FILING DATE: 1999-10-28

PRIOR APPLICATION NUMBER: 60/164,769

PRIOR FILING DATE: 1999-11-12

PRIOR APPLICATION NUMBER: 60/173,383

PRIOR FILING DATE: 1999-12-28

PRIOR APPLICATION NUMBER: 60/175,693

PRIOR FILING DATE: 2000-01-12

PRIOR APPLICATION NUMBER: 60/184,831

PRIOR FILING DATE: 2000-02-24

PRIOR APPLICATION NUMBER: 60/191,637

PRIOR FILING DATE: 2000-03-23

NUMBER OF SEQ ID NOS: 43008

SOFTWARE: FastSeq for windows Version 4.0

SEQ ID NO 34191

LENGTH: 409

TYPE: PRT

ORGANISM: DROSOPHILA

US-11-097-143-34191

Query Match 33.6%; Score 714.5; DB 20; Length 409;

Best Local Similarity 42.0%; Pred. No. 5.7e-61;

Matches 156; Conservative 60; Mismatches 140; Indels 15; Gaps 6;

Qy 39 VLGIETSCDDTAADVDTGAVLGAIHSQTEVHLTKGTGIVPPAAQOLHRENIQRTIVQEA 98

Db 27 VLGIETSCDDTAADVDTGAVLGAIHSQTEVHLTKGTGIVPPAAQOLHRENIQRTIVQEA 86

QY 99 LSASGVSFSDLSAIAATTIKPGLALSLGVLSFSLQVLQOLKKPFIPIHMEHALTRIT 158
DB 87 MEAQLQKPDQLTAIAVTRPGLPLSLVGVRFARHLARLOKPLPIVHMEHALQARE 146
QY 159 N--KVEFPFLVLLISGGHCLLALVQGVSDPFLLGKSLDIAPGMDLKVARRSLIKHPEC 216
DB 147 HPEIGVFPFLCLLASGGHCLVANGPGRLLTIGTLDDBEADDKGRRLILHLPY 206
QY 217 STMSGKAIEHLAK--QGNRFHFDIKRPLHAKNCFSTGLQHTVDKIMKEKEGIEK 275
DB 207 RLWNGRAIEHAQAQADPLAVEFPLPLAQQRNCNCFSTAGIKNNSFRAIRABERARPTP 266
QY 276 GQILSSADIAATVQHTMACHLVKTRTHAILFC--KQDLPQNNAVLVASGVASNFYI 333
DB 267 DGVISNYGDFCAGLRVSRLHMRTORAIEYCLLPHRQLFGDTPPTLVMSGGVANNDAI 326
QY 334 RBAEILTNATQCTLLCPPLCTDNGIMIANNGIERLRAGILHDIE--GIRYEPKCL 392
DB 327 YANIEHLAAQYGCSPFRPSKRYCSQNGVMIAMHGEVL-----LQDKASTRYDYD-- 377
QY 393 GVDISKEVGEA 403
DB 378 SIDIQSAGFA 388
RESULT 14
US-10-424-599-209259
; Sequence 209259, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 209259
; LENGTH: 445
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_3098C.1.pep
US-10-424-599-209259
Query Match 32.1%; Score 681.5; DB 15; Length 445;
Best Local Similarity 43.1%; Pred. No. 1,1e-57;
Matches 155; Conservative 54; Mismatches 126; Indels 25; Gaps 6;
QY 38 IVLGIEFSCDDTAAAVNDEGTGVLGEALHSQTEVHLKTGGIVPPAAQQLHRENIQRIVOE 97
DB 55 VVLGIEFSCDDTAAAVNDEGTGVLGEALHSQTEVHLKTGGIVPPAAQQLHRENIQRIVOE 114
QY 98 ALSASGVSPSLSAIAATTIKPGLALSLGVLSFSLQVLQOLKKPFIPIHMEHALTRIL 157
DB 115 ALDRAVLTEDKLSAVALVITIGPGLSLCLRVGQKAKKAGGFRLLPIIGIHHMEHALVRL 174
QY 158 TNK--VEFPFLVLLISGGHCLLALVQGVSDPFLLGKSLDIAPGMDLKVARRSLIKHPEC 216
DB 175 IEKDLPFPFLMLISGGHCLLALVQGVSDPFLLGKSLDIAPGMDLKVARRSLIKHPEC 231
QY 217 STMSGKAIEHLAKQGNRFHFDIKRPLHAKNCFSTGLQHTVDKIMKEKEGIEK 271
DB 232 --RSGGPAIEHLAKQGNRFHFDIKRPLHAKNCFSTGLQHTVDKIMKEKEGIEK 289
QY 272 GIEKQILSSADIAATVQHTMACHLVKTRTHAILFCQDLPQNNAVLVASGVASNF 331
DB 290 SASNGRDL--SPADIAASFORIAVLHLBERCERAIQWALKMPSIRH--LVVSGVASNQ 345
QY 332 YIRRAEILTNATQCTLLCPPLCTDNGIMIANNGIERLRAGILHDIEGIRYEPKCP 391

DB 346 YIRRAEILTNATQCTLLCPPLCTDNGIMIANNGIERLRAGILHDIEGIRYEPKCP 395
RESULT 15
US-10-067-443-3
; Sequence 3, Application US/10067443
; Publication No. US20030082782A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE HIGHLY EXPRESSED IN
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 3
; LENGTH: 463
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-067-443-3

Query Match 29.8%; Score 634; DB 14; Length 463;
Best Local Similarity 37.9%; Pred. No. 5,6e-53;
Matches 145; Conservative 61; Mismatches 129; Indels 48; Gaps 7;
QY 38 IVLGIEFSCDDTAAAVNDEGTGVLGEALHSQTEVHLKTGGIVPPAAQQLHRENIQRIVOE 97
DB 85 VVLGIEFSCDDTAAAVNDEGTGVLGEALHSQTEVHLKTGGIVPPAAQQLHRENIQRIVOE 141
QY 98 ALSASGVSPSLSAIAATTIKPGLALSLGVLSFSLQVLQOLKKPFIPIHMEHALTRIL 157
DB 142 ALDRAVLTEDKLSAVALVITIGPGLSLCLRVGQKAKKAGGFRLLPIIGIHHMEHALVRL 201
QY 158 T--NKVEFPFLVLLISGGHCLLALVQGVSDPFLLGKSLDIAPGMDLKVARRSLIKHPEC 216
DB 202 VQELSPFPFLMLISGGHCLLALVQGVSDPFLLGKSLDIAPGMDLKVARRSLIKHPEC 258
QY 217 STMSGKAIEHLAKQGNRFHFDIKRPLHAKNCFSTGLQHTVDKIMKEKEGIEK 276
DB 259 --RSGGPAIEHLAKQGNRFHFDIKRPLHAKNCFSTGLQHTVDKIMKEKEGIEK 308
QY 277 QILSSADIAATVQHTMACHLVKTRTHAILFCQDLPQNNAVLVASGVASNFYIRRA 336
DB 309 --IRRRADIAASFORIAVLHLBERCERAIQWALKMPSIRH--LVVSGVASNQ 363
QY 337 LEILTNATQCTLLCPPLCTDNGIMIANNGIERLRAGILHDIEGIRYEPKCP 391
DB 364 LNNIIVENKRLAKLVCPPLCTDNGIMIANNGIERLRAGILHDIEGIRYEPKCP 413
QY 388 -----PKCPGLGVDISKEVGEA 403
DB 414 DYVDLPFRPWPLGEEYAKGRSEA 436

Search completed: November 10, 2005, 16:58:50
Job time : 112.35 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2005, 15:02:11 ; Search time 111.6 Seconds

(without alignment)
1434.756 Million cell updates/sec

Title: US-10-649-273-2

Perfect score: 2125

Sequence: 1 MLITTKTAGVFFPKSKRKVY.....DISKEVGASIKVPLKMEI 414

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_16Dec04:*

1: geneseqp1980s:*
2: geneseqp1980s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2125	100.0	414	5	ABG96478 Novel hum
2	2125	100.0	414	5	ABG96478 Novel hum
3	2125	100.0	414	6	ABG96478 Novel hum
4	2088	98.3	414	6	ABG96478 Novel hum
5	2088	98.3	414	6	ABG96478 Novel hum
6	2088	98.3	414	6	ABG96478 Novel hum
7	1843	85.8	364	6	ADN54471 Human pro
8	1385	65.2	267	5	ABG96487 Novel hum
9	714.5	33.6	409	4	ABG96487 Novel hum
10	659.5	31.0	439	3	ABG96487 Novel hum
11	659.5	31.0	439	3	ABG96487 Novel hum
12	634	29.8	463	3	AAV52216 Arabidops
13	549	25.8	179	5	ABG96489 Novel hum
14	524	24.7	382	6	ABG96489 Novel hum
15	524	24.7	421	5	ABG96489 Novel hum
16	502	23.6	348	6	ABU35232 Protein e
17	501	23.6	350	8	ADL05040 M. catarr
18	499.5	23.5	343	7	ABO68626 Pseudomon
19	494.5	23.3	343	3	ABU39303 Protein e
20	492.5	23.2	341	4	AAU36205 Pseudomon
21	492.5	23.2	341	6	ABU38276 Protein e
22	492.5	23.2	341	7	ADG73342 P aerugin
23	490.5	23.1	341	7	ADG73344 P aerugin
24	489.5	23.0	335	6	ABU27480 Protein e
25	484.5	22.8	337	4	AAU34711 E. coli c

26	484.5	22.8	337	6	ABU47561 Protein e
27	484.5	22.8	337	6	ABU28771 Protein e
28	483.5	22.8	337	6	ABU50237 Protein e
29	482.5	22.7	340	6	ABU40514 Protein e
30	482.5	22.7	357	7	ADP06228 Bacterial
31	481.5	22.7	337	4	AAU38187 Protein e
32	479.5	22.6	337	3	AAV52204 Protein e
33	475.5	22.4	335	2	AAV52203 Protein e
34	475.5	22.4	335	2	AAV52203 Protein e
35	475.5	22.4	342	3	AAV52202 Protein e
36	475.5	22.4	342	4	AAU35450 Haemophil
37	475.5	22.4	342	6	ABU30280 Protein e
38	474.5	22.3	343	7	ABO62704 Klebsiell
39	472.5	22.2	338	6	ABM67812 Phototrab
40	472	22.2	363	7	ADC69682 E. faeciu
41	469	22.1	340	6	ABU24130 Protein e
42	468	22.0	338	6	ABU29893 Protein e
43	465.5	21.9	346	3	AAV52208 Bacillus
44	465.5	21.9	354	6	ABU37844 Protein e
45	465	21.9	341	6	ABU40069 Protein e

ALIGNMENTS

RESULT 1	ABG96478	ABG96478 standard; protein; 414 AA.
ID	ABG96478	
XX	ABG96478;	
AC		
XX		
DT	11-DEC-2002 (first entry)	
XX		
DE	Novel human metalloprotease MPL.	
XX		
XX	Metalloprotease; MP-1; immune disorder; glutamate transport; cancer;	
KW	motor neuron disorder; amyotrophic lateral sclerosis; ALS; diabetes;	
KW	reproductive disorder; Klinefelter's syndrome; germinal cell aplasia;	
KW	genital wart; metabolic disorder; premature puberty; Kallman syndrome;	
KW	Cushing's syndrome; neurodegenerative disease; Alzheimer's disease;	
KW	Parkinson's disease; Huntington's disease; Tourette syndrome; sepsis;	
KW	liver disease; renal disease; immune disorder; rheumatoid arthritis;	
KW	acquired immunodeficiency syndrome; AIDS; pulmonary disease; pneumonia;	
KW	emphysema; cystic fibrosis; vascular disorder; inflammatory disorder;	
XX	neurological disorder.	
XX		
OS	Homo sapiens.	
XX		
PN	W0200272751-A2.	
XX		
PD	19-SEP-2002.	
XX		
PF	05-FEB-2002; 2002WO-US003353.	
XX		
PR	05-FEB-2001; 2001US-0266518P.	
XX		
PR	10-FEB-2001; 2001US-0282814P.	
XX		
PA	(BRIM) BRISTOL-MYERS SQUIBB CO.	
XX		
PI	Chen J, Feder J, Nelson TC, Duclos F, Krystek S;	
XX		
DR	MPL: 2002-723329/78.	
XX		
DR	N-PSDB; ABS76635.	
XX		
PT	New isolated nucleic acid encoding MP-1 protein, useful for preventing,	
XX	treating, or ameliorating diseases associated with aberrant	
PT	metalloprotease activity, e.g. immune, metabolic, inflammatory and	
XX	neurological disorders.	
PS	Claim 5; Fig 1A-C; 473pp; English.	
XX		
CC	The invention describes an isolated nucleic acid molecule (I) encoding a	
XX	metalloprotease (MP-1). (I) is useful for preventing, treating, or	

CC ameliorating a medical condition, particularly an immune disorder, an
 CC aberrant glutamate transport or motor neuron disorder, such as
 CC amyotrophic lateral sclerosis (ALS), its juvenile form or an ALS-like
 CC condition. The compositions and methods are also useful for diagnosing,
 CC prognosticating, treating, ameliorating and/or treating disorders
 CC associated with MP-1 activity, e.g. diabetes, cancer, reproductive
 CC disorders (e.g. Klinefelter's syndrome, genital warts, or germinal cell
 CC aplasia), metabolic disorders (e.g. premature puberty, Kallman syndrome,
 CC or Cushing's syndrome), neurodegenerative diseases (Alzheimer's disease,
 CC Parkinson's disease, Huntington's disease or Tourette syndrome), liver
 CC and renal diseases and immune disorders (e.g. AIDS, rheumatoid arthritis
 CC or sepsis), pulmonary diseases (e.g. pneumonia, emphysema or cystic
 CC fibrosis) and vascular, inflammatory and neurological disorders (e.g.
 CC Alzheimer's disease or Parkinson's disease). This is the amino acid
 CC sequence of a metalloprotease Mpl protein

XX Sequence 414 AA;

Query Match 100.0%; Score 2125; DB 5; Length 414;
 Best Local Similarity 100.0%; Pred. No. 2.2e-215;
 Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLILTKTAGVFPKSKRYEFLRSFNFHPTLFLHKIVLGITSCDDTAAAVDETNV 60
 DB 1 MLILTKTAGVFPKSKRYEFLRSFNFHPTLFLHKIVLGITSCDDTAAAVDETNV 60
 QY 61 LGEAHSQTEVHLKTGIVPPAAQOLHRENIQRIQVEALSASGVSPSDISAIAITTKPGL 120
 DB 61 LGEAHSQTEVHLKTGIVPPAAQOLHRENIQRIQVEALSASGVSPSDISAIAITTKPGL 120
 QY 121 ALSIGVGLSFSIQLVGOLKKPFIPIHMEAHALTITRLTNKVEFPFLVLLISGGHCLALV 180
 DB 121 ALSIGVGLSFSIQLVGOLKKPFIPIHMEAHALTITRLTNKVEFPFLVLLISGGHCLALV 180
 QY 181 QGVSDFLILGKSLDIAPGMDLKVARRSLIKHPECSSTMSGKAIIEHLAKQGRFHPDIK 240
 DB 181 QGVSDFLILGKSLDIAPGMDLKVARRSLIKHPECSSTMSGKAIIEHLAKQGRFHPDIK 240
 QY 241 PRLHAKNCDPSFTGLQHTVDKIIMKKEKEGIEKGQIILSSADIAATVQHTMACHLYKR 300
 DB 241 PRLHAKNCDPSFTGLQHTVDKIIMKKEKEGIEKGQIILSSADIAATVQHTMACHLYKR 300
 QY 301 THRALLFCQKQDILLPQNNAVLVASGGVANSFYIRALIELITNATQCTLLCPPRLCTDNG 360
 DB 301 THRALLFCQKQDILLPQNNAVLVASGGVANSFYIRALIELITNATQCTLLCPPRLCTDNG 360
 QY 361 IMIANGIERLRAGLGIHIDIEGIRYEPKCPGVDSKEVGEASIKVPOLKMEI 414
 DB 361 IMIANGIERLRAGLGIHIDIEGIRYEPKCPGVDSKEVGEASIKVPOLKMEI 414

RESULT 2

AB05481 ID AB05481 standard; protein; 414 AA.

AC AB05481;

DT 19-APR-2002 (first entry)

DE Human O-sialoglycoproteinase-like protein SEQ ID NO:2.

XX Human; O-sialoglycoproteinase-like protein; OSGPLP; enzyme.

XX Homo sapiens.

XX CN1318550-A.

XX 24-OCT-2001.

XX 19-APR-2000; 2000CN-00106834.

XX 19-APR-2000; 2000CN-00106834.

XX

PA (SHAN-) SHANGHAI SHENGYUAN GENE DEV CO LTD.

XX Mao Y, Xie Y;

XX WPI; 2002-115090/16.

XX N-PSDB; ABA93268.

PT O-sialoglycoproteinase-like protein and encoding polynucleotide, useful
 for diagnosing, preventing and treating related diseases.

PS Claim 1; Page 30-31 (Disclosure); 38pp; Chinese.

CC The present sequence represents human O-sialoglycoproteinase-like protein
 CC (OSGPLP). The present invention also describes: (1) the preparation of
 CC the OSGPLP protein; (2) applying the OSGPLP protein in diagnosis; (3) the
 CC prevention and/or treatment of related diseases; (4) utilizing the OSGPLP
 CC protein in screening its agonist, excitomotor and inhibitor and preparing
 CC an antibody against the OSGPLP protein; and (5) the use of the OSGPLP
 CC polynucleotide sequences, proteins, agonists, excitomoters, inhibitors
 CC and antibodies in treating diseases related to the abnormal OSGPLP gene
 CC and in preparing the medicine composite for the treatment

XX Sequence 414 AA;

Query Match 100.0%; Score 2125; DB 5; Length 414;
 Best Local Similarity 100.0%; Pred. No. 2.2e-215;
 Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLILTKTAGVFPKSKRYEFLRSFNFHPTLFLHKIVLGITSCDDTAAAVDETNV 60
 DB 1 MLILTKTAGVFPKSKRYEFLRSFNFHPTLFLHKIVLGITSCDDTAAAVDETNV 60
 QY 61 LGEAHSQTEVHLKTGIVPPAAQOLHRENIQRIQVEALSASGVSPSDISAIAITTKPGL 120
 DB 61 LGEAHSQTEVHLKTGIVPPAAQOLHRENIQRIQVEALSASGVSPSDISAIAITTKPGL 120
 QY 121 ALSIGVGLSFSIQLVGOLKKPFIPIHMEAHALTITRLTNKVEFPFLVLLISGGHCLALV 180
 DB 121 ALSIGVGLSFSIQLVGOLKKPFIPIHMEAHALTITRLTNKVEFPFLVLLISGGHCLALV 180
 QY 181 QGVSDFLILGKSLDIAPGMDLKVARRSLIKHPECSSTMSGKAIIEHLAKQGRFHPDIK 240
 DB 181 QGVSDFLILGKSLDIAPGMDLKVARRSLIKHPECSSTMSGKAIIEHLAKQGRFHPDIK 240
 QY 241 PRLHAKNCDPSFTGLQHTVDKIIMKKEKEGIEKGQIILSSADIAATVQHTMACHLYKR 300
 DB 241 PRLHAKNCDPSFTGLQHTVDKIIMKKEKEGIEKGQIILSSADIAATVQHTMACHLYKR 300
 QY 301 THRALLFCQKQDILLPQNNAVLVASGGVANSFYIRALIELITNATQCTLLCPPRLCTDNG 360
 DB 301 THRALLFCQKQDILLPQNNAVLVASGGVANSFYIRALIELITNATQCTLLCPPRLCTDNG 360
 QY 361 IMIANGIERLRAGLGIHIDIEGIRYEPKCPGVDSKEVGEASIKVPOLKMEI 414
 DB 361 IMIANGIERLRAGLGIHIDIEGIRYEPKCPGVDSKEVGEASIKVPOLKMEI 414

RESULT 3

ABJ26654 ID ABJ26654 standard; protein; 414 AA.

AC ABJ26654;

DT 01-MAY-2003 (first entry)

DE Human protein modification + maintenance molecule protein SEQ ID NO:8.

XX Cystostatic; antiarteriosclerotic; anti-HIV; antiallergic; anticonvulsant;

XX cerebroprotective; antiparkinsonian; nootropic; antiinflammatory;

XX anticancer; hepatocellular; gynaecological; antibacterial; virucide;

XX protozoacide; antiparasitic; cell proliferative disease; PMOD;

XX protein modification and maintenance molecule; immunogenic fragment;

XX cancer; autoimmune; inflammatory disease; neurological disorder;

PN WO200274960-A2.
 XX 26-SEP-2002.
 XX 08-NOV-2001; 2001WO-US051427.
 XX 08-NOV-2000; 2000US-0246768P.
 PR 08-NOV-2000; 2000US-0246772P-yes
 PR 15-NOV-2000; 2000US-0249185P.
 XX (MILL-) MILLENNIUM PHARM INC.
 PA
 PI Leyby KR, Kapeller-Libermann R, Glucksmann M;
 XX WPI: 2002-759898/82.
 DR N-PSDB; AAD46856.
 PT New 38650, 28472, 5495, 65507, 81588 or 14354 nucleic acid molecules,
 PT useful for diagnosing and treating cancer, immune, cardiovascular,
 PT hematopoietic, brain, pain, metabolic, liver or platelet disorders, and
 PT in pharmacogenomics.
 XX
 PS Claim 1; Fig 8; 178pp; English.
 CC The present invention relates to novel 38650, 28472, 5495, 65507, 81588
 CC or 14354 nucleic acid molecules which encode adenosine deaminase, glyco-
 CC protease or seven transmembrane domain (7TM) receptor family members.
 CC Sequences of seven transmembrane domain (7TM) receptor family members.
 CC or aberrant cellular proliferation and/or differentiation (e.g. colon or
 CC lung cancer), immune disorders (e.g. selective IGA deficiency, rheumatoid
 CC arthritis, diabetes, allergy), heart or cardiovascular disorders (e.g.,
 CC hypertension, atherosclerosis, arrhythmias, ischaemic heart disease,
 CC myocardial infarction, thrombus) including endothelial cell disorders
 CC (e.g. psoriasis, Grave's disease), haematopoietic disorders, brain
 CC pain and metabolic disorders (e.g. obesity), Parkinson's or Alzheimer's disease),
 CC disorders. They are also useful in screening assays, predictive medicine
 CC (e.g. diagnostic assays, prognostic assays, monitoring clinical trials
 CC and pharmacogenetics) and prophylactic and therapeutic methods. The
 CC nucleic acids may also be used in chromosome mapping, tissue typing and
 CC forensic biology and as surrogate markers. Sequences of the invention are
 CC also used in gene therapy. The present sequence is human glycoprotease
 CC 28472 protein
 XX
 XX Sequence 414 AA;

Query Match 98.3%; Score 2088; DB 5; Length 414;
 Best Local Similarity 98.3%; Pred. No. 1.9e-211;
 Matches 407; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 MLITVTAGVFFPKSRKRYEPLRSFNPFGTLFLHKIVIGIETSCDDTAADVDEGNV 60
 DB 1 MLITVTAGVFFPKSRKRYEPLRSFNPFGTLFLHKIVIGIETSCDDTAADVDEGNV 60
 QY 61 LGAHISQREVLHKTGQIVPPAAQOLHRENIQRIVOALASGVSPPSLAIAATTIRGL 120
 DB 61 LGAHISQREVLHKTGQIVPPAAQOLHRENIQRIVOALASGVSPPSLAIAATTIRGL 120
 QY 121 ALSIGVGLSFSLQVGVOLKKKPIPIHMEAHALITRLTNKVEFFVLILISGHCILALV 180
 DB 121 ALSIGVGLSFSLQVGVOLKKKPIPIHMEAHALITRLTNKVEFFVLILISGHCILALV 180
 QY 181 QGVSDFLILGKSLDIAFGDMLDKVARLSIIKHPECSTMSGKAIIEHLAKGNFPHDIK 240
 DB 181 QGVSDFLILGKSLDIAFGDMLDKVARLSIIKHPECSTMSGKAIIEHLAKGNFPHDIK 240
 QY 241 PPIHAAKNCDFSTGLOHTVDTKIIIMKEKEBGEIKGQILSSAADIATVQHTMACHVVR 300
 DB 241 PPIHAAKNCDFSTGLOHTVDTKIIIMKEKEBGEIKGQILSSAADIATVQHTMACHVVR 300
 QY 301 THRAILFCQKRDILLPNNNAVIVASGVASNPYIRRALEITLNATQCTLLCPPLCTDNG 360
 DB 301 THRAILFCQKRDILLPNNNAVIVASGVASNPYIRRALEITLNATQCTLLCPPLCTDNG 360

QY 361 IMIANGIERLRAGLGIHDIBGIRYEPKCPGLGVDISKVEGEASIKVPOLMEI 414
 DB 361 IMIANGIERLRAGLGIHDIBGIRYEPKCPGLGVDISKVEGEASIKVPOLMEI 414
 RESULT 5
 ABG71161
 ID ABG71161 standard; protein; 414 AA.
 XX
 AC ABG71161;
 XX
 DT 30-JAN-2003 (first entry)
 XX
 DE Novel human glycoprotease 28472.
 XX
 KW Cancer; aberrant cell proliferation; aberrant cell differentiation;
 KW breast cancer; ovarian cancer; prostate cancer; colon cancer;
 KW lung cancer; immune disorder; heart disorder; cardiovascular disorder;
 KW endothelial disorder; hematopoietic disorder; blood vessel disorder;
 KW brain disorder; pain; metabolic disorder; liver disorder; diabetes;
 KW platelet disorder; carcinoma; sarcoma; leukaemia; Hodgkin's disease;
 KW autoimmune disorder; hypertension; atherosclerosis; heart failure;
 KW myocardial infarction; ischaemic heart disease; Crohn's disease;
 KW Grave's disease; Kawasaki syndrome; Raynaud's disease; aneurysm;
 KW cerebral ischaemia; peripheral neuropathy; Alzheimer's disease;
 KW Parkinson's disease; anorexia nervosa; cachexia; glycoprotease; enzyme.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Domain 138..152
 FT /label= Glycoprotease_domain
 XX
 XX WO200277233-A2.
 FT 03-OCT-2002.
 XX
 PD 08-NOV-2001; 2001WO-US046724.
 XX
 PP 08-NOV-2001; 2001WO-US046724.
 XX
 PR 08-NOV-2000; 2000US-0246768P-yes
 PR 08-NOV-2000; 2000US-0246772P-yes
 PR 15-NOV-2000; 2000US-0249185P.
 XX
 PA (MILL-) MILLENNIUM PHARM INC.
 XX
 PI Leyby KR, Kapeller-Libermann R, Glucksmann M;
 XX WPI: 2003-029938/02.
 DR N-PSDB; ABS57020.
 XX
 PT New adenosine deaminase, glycoprotease and seven transmembrane domain
 PT nucleic acids and polypeptides, designated 38650, 28472, 5495, 65507,
 PT 81588 and 14354, useful for treating e.g. leukemias, Hodgkin's disease or
 PT hypertension.
 XX
 PS Claim 4; Fig 8A-B; 178pp; English.
 XX
 CC The invention describes isolated 38650, 28472, 5495, 65507, 81588 and
 CC 14354 nucleic acid molecules (I) and their encoded polypeptides (II). The
 CC 38650 nucleic acid molecule comprises a sequence encoding adenosine
 CC deaminase. The 28472 nucleic acid molecule comprises a sequence encoding
 CC a human glycoprotease. The 5495, 65507, 81588 and 14354 comprise
 CC sequences that encode a human seven transmembrane domain (7TM). The
 CC 38650, 28472, 5495, 65507, 81588 or 14354 nucleic acid and polypeptide
 CC sequences are useful for diagnosing, preventing or treating a subject
 CC with or at risk of developing a disorder, e.g. cancer or aberrant
 CC cellular proliferation and/or differentiation (e.g. breast, ovarian,
 CC prostate, colon or lung cancer), immune disorders, heart disorders,
 CC cardiovascular disorders, endothelial disorders, hematopoietic disorders,
 CC blood vessel disorders, brain disorders, pain and metabolic disorders,
 CC liver disorders or platelet disorders. These disorders include carcinoma,
 CC sarcoma, leukaemia, Hodgkin's disease, autoimmune disorders,

CC hypertension, atherosclerosis, heart failure, myocardial infarction,
 CC ischaemic heart disease, Crohn's disease, Grave's disease, Kawasaki
 CC syndrome, Raynaud's disease, aneurysm, cerebral ischaemia, peripheral
 CC neuropathy, Alzheimer's disease, Parkinson's disease, anorexia nervosa,
 CC cachexia or diabetes. This is the amino acid sequence of the novel human
 CC glycoprotease 28472
 XX
 SQ Sequence 414 AA;
 Query Match 98.3%; Score 2088; DB 6; Length 414;
 Best Local Similarity 98.3%; Pred. No. 1.9e-211;
 Matches 407; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
 QY 1 MLILTKTAGVFFPKSRKRYEFLRSFNFHGTLLFLHKIVLGITSCDDTAAAVDEGNY 60
 DB 1 MLILTKTAGVFFPKSRKRYEFLRSFNFHGTLLFLHKIVLGITSCDDTAAAVDEGNY 60
 QY 61 LGEAHSQTEVHLKTGIVPPAAQOLHRENIQRIQVEALSASGVSPDSLAIATTIKPGL 120
 DB 61 LGEAHSQTEVHLKTGIVPPAAQOLHRENIQRIQVEALSASGVSPDSLAIATTIKPGL 120
 QY 121 ALSIGVLSFSLQVLGQKKPFIPIHMEAHALTIRLTNKKVEPFLVLLISGHCILLAV 180
 DB 121 ALSIGVLSFSLQVLGQKKPFIPIHMEAHALTIRLTNKKVEPFLVLLISGHCILLAV 180
 QY 181 QGVSDFLILGKSLDIAPGMDLVKVARSLIKHPECSSTSGGKAIETHLAKQGRFFFDIK 240
 DB 181 QGVSDFLILGKSLDIAPGMDLVKVARSLIKHPECSSTSGGKAIETHLAKQGRFFFDIK 240
 QY 241 PPLHAKNCDPSFTGLQHTVDKTIIMKKEKEGIEKQIILSSADIAATVQHTMACHLVKR 300
 DB 241 PPLHAKNCDPSFTGLQHTVDKTIIMKKEKEGIEKQIILSSADIAATVQHTMACHLVKR 300
 QY 301 THRALLFCQRDLTPONNAVIVASGVASNFYIRALLETITNAQTCTLLCPPRLCTDNG 360
 DB 301 THRALLFCQRDLTPONNAVIVASGVASNFYIRALLETITNAQTCTLLCPPRLCTDNG 360
 QY 361 IMIANNGIERLRAGILHDIGIRYEPKCPGLVDISKVEGASIVPOLKMEI 414
 DB 361 IMIANNGIERLRAGILHDIGIRYEPKCPGLVDISKVEGASIVPOLKMEI 414
 RESULT 6
 ABU09569 standard; protein; 414 AA.
 AC ABU09569;
 XX
 DT 08-JUL-2003 (first entry)
 XX
 DE Human glycoprotease encoded by cDNA 28472.
 XX
 KW Human; enzyme; cancer; aberrant cellular proliferation; differentiation;
 KW immune disorders; heart disorder; brain disorder;
 KW cardiovascular disorder; endothelial cell disorder; pain disorder;
 KW haematopoietic disorder; blood vessel disorder; metabolic disorder;
 KW liver disorder; platelet disorder; glycoprotease.
 XX
 OS Homo sapiens.
 XX
 PN US2003009017-A1.
 XX
 PD 09-JAN-2003.
 XX
 XX 08-NOV-2001; 2001US-00012140.
 XX
 XX 08-NOV-2000; 2000US-0246768P.
 PR 08-NOV-2000; 2000US-0246772P. **YES**
 PR 15-NOV-2000; 2000US-0249185P.
 XX
 XX (LEIBY/) LEIBY K R.
 PA (KAPE/) KAPLELLER-LIBERMANN R.
 PA (GLUC/) GLUCKSMANN M A.

XX
 PT Leiby KR, Kapeller-Libermann R, Glucksmann MA;
 XX
 XX WPI; 2003-42888/40.
 DR N-PSDB; ACA60887.
 XX
 PT New isolated 3650, 28472, 5495, 65507, 81588 and 14354 nucleic acid
 PT molecules, useful for diagnosing, treating cancer, pain, or immune,
 PT heart, endothelial cell, haematopoietic, blood vessel, brain, metabolic
 PT and liver disorders.
 XX
 PS Claim 4; Fig 8; 90pp; English.
 XX
 CC The invention relates to an isolated 3650 (encoding adenosine
 CC deaminase), 28472 (encoding a glycoprotease), 5495 (encoding a 7
 CC transmembrane receptor), 65507, 81588 or 14354 nucleic acid molecule (or
 CC a sequence which is at least 60% identical to the six nucleic acids or
 CC their open reading frames, fragments of at least 15 nucleotides,
 CC naturally occurring variants, or a DNA insert of the plasmid deposited
 CC with the American Type Culture Collection as Accession No. not defined in
 CC the specification, which encodes the amino acid sequence). Also included
 CC are a host cell containing the nucleic acids (used to produce the
 CC proteins), the encoded proteins, an antibody that selectively binds to
 CC the polypeptide, and identifying a compound that binds to/modulates the
 CC activity of the polypeptide. The nucleic acid molecules, polypeptides and
 CC methods are useful for diagnosing, treating cancer, aberrant cellular
 CC proliferation and/or differentiation, immune disorders, heart disorders,
 CC cardiovascular disorders including endothelial cell disorders,
 CC haematopoietic disorders, blood vessel disorders, brain disorders, pain
 CC and metabolic disorders, liver disorders and platelet disorders (many
 CC examples of these disorders are given in the specification). The present
 CC sequence is the Human glycoprotease encoded by cDNA 28472
 XX
 SQ Sequence 414 AA;
 Query Match 98.3%; Score 2088; DB 6; Length 414;
 Best Local Similarity 98.3%; Pred. No. 1.9e-211;
 Matches 407; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
 QY 1 MLILTKTAGVFFPKSRKRYEFLRSFNFHGTLLFLHKIVLGITSCDDTAAAVDEGNY 60
 DB 1 MLILTKTAGVFFPKSRKRYEFLRSFNFHGTLLFLHKIVLGITSCDDTAAAVDEGNY 60
 QY 61 LGEAHSQTEVHLKTGIVPPAAQOLHRENIQRIQVEALSASGVSPDSLAIATTIKPGL 120
 DB 61 LGEAHSQTEVHLKTGIVPPAAQOLHRENIQRIQVEALSASGVSPDSLAIATTIKPGL 120
 QY 121 ALSIGVLSFSLQVLGQKKPFIPIHMEAHALTIRLTNKKVEPFLVLLISGHCILLAV 180
 DB 121 ALSIGVLSFSLQVLGQKKPFIPIHMEAHALTIRLTNKKVEPFLVLLISGHCILLAV 180
 QY 181 QGVSDFLILGKSLDIAPGMDLVKVARSLIKHPECSSTSGGKAIETHLAKQGRFFFDIK 240
 DB 181 QGVSDFLILGKSLDIAPGMDLVKVARSLIKHPECSSTSGGKAIETHLAKQGRFFFDIK 240
 QY 241 PPLHAKNCDPSFTGLQHTVDKTIIMKKEKEGIEKQIILSSADIAATVQHTMACHLVKR 300
 DB 241 PPLHAKNCDPSFTGLQHTVDKTIIMKKEKEGIEKQIILSSADIAATVQHTMACHLVKR 300
 QY 301 THRALLFCQRDLTPONNAVIVASGVASNFYIRALLETITNAQTCTLLCPPRLCTDNG 360
 DB 301 THRALLFCQRDLTPONNAVIVASGVASNFYIRALLETITNAQTCTLLCPPRLCTDNG 360
 QY 361 IMIANNGIERLRAGILHDIGIRYEPKCPGLVDISKVEGASIVPOLKMEI 414
 DB 361 IMIANNGIERLRAGILHDIGIRYEPKCPGLVDISKVEGASIVPOLKMEI 414
 RESULT 7
 ADA54471 standard; protein; 364 AA.
 ID ADA54471
 XX
 AC ADA54471;

XX 20-NOV-2003 (first entry)
 XX Human protein, SEQ ID 2039.
 DE
 XX Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;
 KM Gene therapy; human; secretory protein; membrane proteins; cancer;
 KM inflammatory disease; osteoporosis; neurological disease.
 XX
 OS Homo sapiens.
 XX
 PN EPI293569-A2.
 XX
 PD 19-MAR-2003.
 XX
 PF 21-MAR-2002; 2002EP-00006586.
 XX
 XX 14-SEP-2001; 2001JP-00328381.
 PR 24-JAN-2002; 2002US-0350435P.
 XX
 PA (HELI-) HELIX RES INST.
 XX (REAS-) RES ASSOC BIOTECHNOLOGY.
 XX
 PI Isegai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
 PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
 PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuno Y;
 XX
 DR WPI; 2003-395539/38.
 DR N-PSDB; ADA52832.
 XX
 PT New polynucleotides encoding full-length polypeptides, e.g. secretory
 PT and/or membrane proteins, useful for developing medicines for diseases in
 PT which the gene is involved, or as target molecules for gene therapy.
 XX
 PS Claim 14; SEQ ID NO 2039; 205pp; English.
 XX
 CC The present invention relates to novel human secretory or membrane
 CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
 CC ADA54071). The coding sequences are useful in the gene therapy of
 CC diseases caused by abnormalities of the proteins, e.g. cancer,
 CC inflammatory diseases, osteoporosis or neurological disease.
 CC
 XX
 SQ Sequence 364 AA;
 Query Match 86.8%; Score 1845; DB 6; Length 364;
 Best Local Similarity 99.2%; Pred. No. 7,9e-186;
 Matches 361; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MLILTKTGVFPFKPSRKRYEFLRSFNPFGTLFLHKIYLGIEETSCDDTAATAAVDETGNV 60
 DB 1 MLILTKTGVFPFKPSRKRYEFLRSFNPFGTLFLHKIYLGIEETSCDDTAATAAVDETGNV 60
 QY 61 LGEAHSOTVEHLKTGGIVPPAAQOLHRENIORIVQEALASAGVSPDSALATTTKPGI 120
 DB 61 LGEAHSOTVEHLKTGGIVPPAAQOLHRENIORIVQEALASAGVSPDSALATTTKPGI 120
 QY 121 ALSIGVGVSFSLQVLQGLKKPPIPIHMEAHALTITLTKVPEPPVLLISGCHLLALV 180
 DB 121 ALSIGVGVSFSLQVLQGLKKPPIPIHMEAHALTITLTKVPEPPVLLISGCHLLALV 180
 QY 121 ALSLGVGVSFSLQVLQGLKKPPIPIHMEAHALTITLTKVPEPPVLLISGCHLLALV 180
 DB 121 ALSLGVGVSFSLQVLQGLKKPPIPIHMEAHALTITLTKVPEPPVLLISGCHLLALV 180
 QY 181 QGVSPFLILGKSLDIPAGMDLKVARRLLIHPBSTMSGGAIETHLAKQGRFHPDIK 240
 DB 181 QGVSPFLILGKSLDIPAGMDLKVARRLLIHPBSTMSGGAIETHLAKQGRFHPDIK 240
 QY 241 PPIHAKKNCDFSTGTQHTYTDKIIMKEKEBEGIEKQIISASADIATVQHTACHLVKR 300
 DB 241 PPIHAKKNCDFSTGTQHTYTDKIIMKEKEBEGIEKQIISASADIATVQHTACHLVKR 300
 QY 301 THRALLFCQRDLDPONNAVLVAGGVANFYIRRALILTNATQCTLLCPPRLCTDNG 360
 DB 301 THRALLFCQRDLDPONNAVLVAGGVANFYIRRALILTNATQCTLLCPPRLCTDNG 360
 QY 361 IMIA 364

DB 361 IMIA 364
 RESULT 8
 ID ABG96487 standard; protein; 267 AA.
 AC ABG96487;
 XX
 DT 11-DEC-2002 (first entry)
 XX
 DE Novel human metalloprotease MPI fragment #1.
 XX
 KM Metalloprotease; MP-1; immune disorder; glutamate transport; cancer;
 KM motor neuron disorder; amyotrophic lateral sclerosis; ALS; diabetes;
 KM reproductive disorder; Klinefelter's syndrome; germinal cell aplasia;
 KM genital wart; metabolic disorder; premature puberty; Kallman syndrome;
 KM Cushing's syndrome; neurodegenerative disease; Alzheimer's disease;
 KM Parkinson's disease; Huntington's disease; Tourette syndrome; sepsis;
 KM liver disease; renal disease; immune disorder; rheumatoid arthritis;
 KM acquired immunodeficiency syndrome; AIDS; pulmonary disease; pneumonia;
 KM emphysema; cystic fibrosis; vascular disorder; inflammatory disorder;
 KM neurological disorder.
 XX
 OS Homo sapiens.
 XX
 PN MO200272751-A2.
 XX
 PD 19-SEP-2002.
 XX
 PF 05-FEB-2002; 2002WO-US003353.
 XX
 PR 05-FEB-2001; 2001US-0266518P.
 PR 10-APR-2001; 2001US-0282814P.
 XX
 XX (BRIM) BRISTOL-MYERS SQUIBB CO.
 XX
 PI Chen J, Feder J, Nelson TC, Duclos F, Krystek S;
 PI
 DR WPI; 2002-723329/78.
 DR N-PSDB; ABS76639.
 XX
 PT New isolated nucleic acid encoding MP-1 protein, useful for preventing,
 PT treating, or ameliorating diseases associated with aberrant
 PT metalloproteinase activity, e.g. immune, metabolic, inflammatory and
 PT neurological disorders.
 XX
 PS Claim 5; Page 29; 473pp; English.
 XX
 CC The invention describes an isolated nucleic acid molecule (I) encoding a
 CC metalloprotease (MP-1). (I) is useful for preventing, treating, or
 CC ameliorating a medical condition, particularly an immune disorder, an
 CC aberrant glutamate transport or motor neuron disorder, such as
 CC amyotrophic lateral sclerosis (ALS), its juvenile form or an ALS-like
 CC condition. The compositions and methods are also useful for diagnosing,
 CC prognosticating, treating, ameliorating and/or treating disorders
 CC associated with MP-1 activity, e.g. diabetes, cancer, reproductive
 CC disorders (e.g. Klinefelter's syndrome, genital warts, or germinal cell
 CC aplasia), metabolic disorders (e.g. premature puberty, Kallman syndrome,
 CC or Cushing's syndrome), neurodegenerative diseases (Alzheimer's disease,
 CC Parkinson's disease, Huntington's disease or Tourette syndrome), liver
 CC and renal diseases and immune disorders (e.g. AIDS, rheumatoid arthritis
 CC or sepsis), pulmonary diseases (e.g. pneumonia, emphysema or cystic
 CC fibrosis) and vascular, inflammatory and neurological disorders (e.g.
 CC Alzheimer's disease or Parkinson's disease). This is the amino acid
 CC sequence of a metalloprotease MPI protein
 XX
 SQ Sequence 267 AA;
 Query Match 65.2%; Score 1385; DB 5; Length 267;
 Best Local Similarity 100.0%; Pred. No. 2,4e-137;
 Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

PR	06-MAY-1999;	99US-0132486P.	PR	22-JUL-1999;	99US-0145089P.
PR	07-MAY-1999;	99US-0132487P.	PR	22-JUL-1999;	99US-0145129P.
PR	06-MAY-1999;	99US-0132863P.	PR	23-JUL-1999;	99US-0145145P.
PR	11-MAY-1999;	99US-0134256P.	PR	23-JUL-1999;	99US-0145218P.
PR	14-MAY-1999;	99US-0134218P.	PR	23-JUL-1999;	99US-0145224P.
PR	14-MAY-1999;	99US-0134219P.	PR	26-JUL-1999;	99US-0145276P.
PR	14-MAY-1999;	99US-0134221P.	PR	27-JUL-1999;	99US-0145913P.
PR	14-MAY-1999;	99US-0134221P.	PR	27-JUL-1999;	99US-0145918P.
PR	18-MAY-1999;	99US-0134768P.	PR	27-JUL-1999;	99US-0145951P.
PR	19-MAY-1999;	99US-0134941P.	PR	28-JUL-1999;	99US-0145951P.
PR	20-MAY-1999;	99US-0135124P.	PR	02-AUG-1999;	99US-0146386P.
PR	21-MAY-1999;	99US-0135353P.	PR	02-AUG-1999;	99US-0146388P.
PR	24-MAY-1999;	99US-0135629P.	PR	03-AUG-1999;	99US-0146389P.
PR	25-MAY-1999;	99US-0136021P.	PR	03-AUG-1999;	99US-0147038P.
PR	27-MAY-1999;	99US-0136392P.	PR	04-AUG-1999;	99US-0147204P.
PR	28-MAY-1999;	99US-0136782P.	PR	04-AUG-1999;	99US-0147302P.
PR	01-JUN-1999;	99US-0137222P.	PR	05-AUG-1999;	99US-0147192P.
PR	03-JUN-1999;	99US-0137528P.	PR	05-AUG-1999;	99US-0147260P.
PR	04-JUN-1999;	99US-0137502P.	PR	06-AUG-1999;	99US-0147303P.
PR	07-JUN-1999;	99US-0137724P.	PR	06-AUG-1999;	99US-0147416P.
PR	08-JUN-1999;	99US-0138094P.	PR	09-AUG-1999;	99US-0147493P.
PR	10-JUN-1999;	99US-0138540P.	PR	09-AUG-1999;	99US-0147935P.
PR	10-JUN-1999;	99US-0138847P.	PR	10-AUG-1999;	99US-0148171P.
PR	14-JUN-1999;	99US-0139119P.	PR	11-AUG-1999;	99US-0148311P.
PR	16-JUN-1999;	99US-0139452P.	PR	12-AUG-1999;	99US-0148341P.
PR	16-JUN-1999;	99US-0139453P.	PR	13-AUG-1999;	99US-0148565P.
PR	17-JUN-1999;	99US-0139453P.	PR	13-AUG-1999;	99US-0148684P.
PR	18-JUN-1999;	99US-01394592P.	PR	16-AUG-1999;	99US-01493368P.
PR	18-JUN-1999;	99US-0139454P.	PR	17-AUG-1999;	99US-0149175P.
PR	18-JUN-1999;	99US-0139455P.	PR	18-AUG-1999;	99US-0149426P.
PR	18-JUN-1999;	99US-0139456P.	PR	20-AUG-1999;	99US-0149722P.
PR	18-JUN-1999;	99US-0139457P.	PR	20-AUG-1999;	99US-0149723P.
PR	18-JUN-1999;	99US-0139458P.	PR	20-AUG-1999;	99US-0149929P.
PR	18-JUN-1999;	99US-0139459P.	PR	23-AUG-1999;	99US-0149902P.
PR	18-JUN-1999;	99US-0139460P.	PR	23-AUG-1999;	99US-0149930P.
PR	18-JUN-1999;	99US-0139461P.	PR	25-AUG-1999;	99US-0150566P.
PR	18-JUN-1999;	99US-0139462P.	PR	26-AUG-1999;	99US-0150884P.
PR	18-JUN-1999;	99US-0139463P.	PR	27-AUG-1999;	99US-0151065P.
PR	18-JUN-1999;	99US-0139763P.	PR	27-AUG-1999;	99US-0151066P.
PR	18-JUN-1999;	99US-0139763P.	PR	27-AUG-1999;	99US-0151083P.
PR	21-JUN-1999;	99US-0139817P.	PR	30-AUG-1999;	99US-0151303P.
PR	22-JUN-1999;	99US-0139899P.	PR	31-AUG-1999;	99US-0151438P.
PR	23-JUN-1999;	99US-0140353P.	PR	01-SEP-1999;	99US-0151930P.
PR	23-JUN-1999;	99US-0140354P.	PR	07-SEP-1999;	99US-0152263P.
PR	24-JUN-1999;	99US-0140695P.	PR	10-SEP-1999;	99US-0153070P.
PR	28-JUN-1999;	99US-0140823P.	PR	13-SEP-1999;	99US-0153758P.
PR	29-JUN-1999;	99US-0140981P.	PR	15-SEP-1999;	99US-0154018P.
PR	30-JUN-1999;	99US-0141287P.	PR	16-SEP-1999;	99US-0154039P.
PR	01-JUL-1999;	99US-0141842P.	PR	20-SEP-1999;	99US-0154779P.
PR	01-JUL-1999;	99US-0142154P.	PR	22-SEP-1999;	99US-0155139P.
PR	02-JUL-1999;	99US-0142055P.	PR	23-SEP-1999;	99US-0155186P.
PR	06-JUL-1999;	99US-0142390P.	PR	24-SEP-1999;	99US-0155659P.
PR	08-JUL-1999;	99US-0142803P			

XX Arabidopsis thaliana yjld protein homologue.
 XX yjld protein; essential; Gram positive; Gram negative; conserved; motif;
 KW identification; antagonist; antibacterial; antibiotic; broad spectrum;
 KW treatment; infection; resistance; drug target.
 OS Arabidopsis thaliana.
 XX
 XX Key Location/Qualifiers
 FT Region 86..96
 FT /note="yjld conserved motif 3"
 FT Region 111..131
 FT /note="yjld conserved motif 4"
 FT Region 152..198
 FT /note="yjld conserved motif 2"
 FT Region 208..259
 FT /note="yjld conserved motif 1"
 FT
 FN WO954470-A2.
 PD
 PD 28-OCT-1999.
 PF 20-APR-1999; 99WO-EP002635.
 PR 22-APR-1998; 98GB-00008423.
 PA (GLAX) GLAXO GROUP LTD.
 PI Arigoni F, Edgerton MD, Loferer H, Peitsch MC,
 DR WPI; 2000-013253/01.
 PT Novel bacterial polypeptides used to identify broad spectrum antibiotics.
 PS
 PS Claim 1; Fig 1; 55pp; English.
 CC Proteins AAY52202-Y52217 encompass a novel family of proteins designated
 CC the yjld family, after the name given to the Escherichia coli family
 CC member. These proteins are essential for the survival of both Gram
 CC negative and Gram positive bacteria, although no function has as yet been
 CC ascribed to these proteins. The yjld proteins, fragments of yjld proteins
 CC (for example, fragments encompassing one or more conserved yjld motifs
 CC such as AAY52218-Y52284) and nucleotides encoding them can be used to
 CC identify antagonists and broad spectrum antibacterial compounds. These
 CC antagonists and compounds can be used to treat a wide range of bacterial
 CC infections. New antibiotics are urgently needed, as serious bacterial
 CC infections and antibiotic resistant strains are becoming increasingly
 CC prevalent. The proteins of the invention are essential proteins for
 CC bacterial viability, and represent new targets for antibiotics
 CC
 SO Sequence 463 AA;
 Query Match 29.8%; Score 634; DB 3; Length 463;
 Best Local Similarity 37.9%; Pred. No. 1.8e-57;
 Matches 145; Conservative 61; Mismatches 129; Indels 48; Gaps 7;
 Oy 38 IVLGIETSCDDTAAAVVDETGVNLGSAHISQTEVHLKTCGIYPPAAQOLHRENTQRIYD 97
 Db 85 VVLGIETSCDDTAAAVVSPFNHLSSEC---RAELLVQYGVGVAIPKQAEASHRVIDYKVD 141
 Oy 98 ALSASGVSPDSLSATATTKPGALSLIGVLSFSLQVLQKKPFIPIHMMEXHATITL 157
 Db 142 ALDKNLTLEKDISSAAVAVITIGELISCLRVGVKARVAANFSPITVGVHMMEXHAAVLAR 201
 Oy 158 T-NKVEPPLVLLISGHCILALVGVSDPLLLKSLIDAPGMDLKVARRSLIGKHP 216
 Db 202 VVGELISFPMALLISGHCNLLVLAHKLCQYTOLGTTVDALIGAPFKTKAMGLDMH--- 258
 Oy 217 STMGSKAKI RHT AKQGNPFPII KQPLAHNAKXCDSPGTLGNTDGI HKKKEKEGIEKG 276
 Db 259 --RSGGPAVEELALBGDAKSVKFNVMKHKDCNFSYAGLKTQVRLAIEAKG----- 308

Oy 277 QILSSADIDATYQHTMACHLVKRTTRAILPCKORDLLPOMNAVLYVSGVASFYRRA 336
 Db 309 --IRNRADIASFORAVLHLEKCEKRAIDWALE---LEPSIKMTVIGGVASNKYRLR 363
 Oy 337 LEILITNAQCTLLCPPLRLCTDNGIMAMNGIERLAGLILHDIEGIRYE----- 387
 Db 364 LNNIVENKULKLVCPPLSLCTDNGVAMVMTGLEHFRVG-----RIDPPPAPE 413
 Oy 388 -----PKCPLGVDSKEVGEA 403
 Db 414 DYYVDLRPRMPLGEYAKGRSEA 436
 RESULT 13
 ABG96489
 ID ABG96489 standard; protein; 179 AA.
 XX
 AC ABG96489;
 XX
 DT 11-DEC-2002 (first entry)
 XX
 DE Novel human metalloprotease MPI fragment #3.
 XX
 KW Metalloprotease; MP-1; immune disorder; glutamate transport; cancer;
 KW motor neuron disorder; amyotrophic lateral sclerosis; ALS; diabetes;
 KW reproductive disorder; Klinefelter's syndrome; germinal cell aplasia;
 KW genital wart; metabolic disorder; premature puberty; Kallman syndrome;
 KW Cushing's syndrome; neurodegenerative disease; Alzheimer's disease;
 KW Parkinson's disease; Huntington's disease; Tourette syndrome; sepsis;
 KW liver disease; renal disease; immune disorder; rheumatoid arthritis;
 KW acquired immunodeficiency syndrome; AIDS; pulmonary disease; pneumonia;
 KW emphysema; cystic fibrosis; vascular disorder; inflammatory disorder;
 KW neurological disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200272751-A2.
 PD 19-SEP-2002.
 PF 05-FEB-2002; 2002WO-US003353.
 PR 05-FEB-2001; 2001US-0266518P.
 PA 10-APR-2001; 2001US-0282814P.
 XX (BRIM) BRISTOL-MYERS SQUIBB CO.
 PI Chen J, Feder J, Nelson TC, Duclos F, Krystek S;
 DR WPI; 2002-723329/78.
 PT New isolated nucleic acid encoding MP-1 protein, useful for preventing,
 PT treating, or ameliorating diseases associated with aberrant
 PT metalloprotease activity, e.g. immune, metabolic, inflammatory and
 PT neurological disorders.
 XX
 PS Disclosure; Page 50; 473pp; English.
 XX
 CC The invention describes an isolated nucleic acid molecule (I) encoding a
 CC metalloprotease (MP-1). (I) is useful for preventing, treating, or
 CC ameliorating a medical condition, particularly an immune disorder, an
 CC aberrant glutamate transport or motor neuron disorder, such as
 CC amyotrophic lateral sclerosis (ALS), its juvenile form or an ALS-like
 CC condition. The compositions and methods are also useful for diagnosing,
 CC prognosticating, treating, ameliorating and/or treating disorders
 CC associated with MP-1 activity, e.g. diabetes, cancer, reproductive
 CC disorders (e.g. Klinefelter's syndrome, genital warts, or germinal cell
 CC aplasia), metabolic disorders (e.g. premature puberty, Kallman syndrome,
 CC or Cushing's syndrome), neurodegenerative diseases (Alzheimer's disease,
 CC Parkinson's disease, Huntington's disease or Tourette syndrome), liver
 CC and renal diseases and immune disorders (e.g. AIDS, rheumatoid arthritis
 CC or sepsis), pulmonary diseases (e.g. pneumonia, emphysema or cystic
 CC fibrosis) and vascular, inflammatory and neurological disorders (e.g.

CC Alzheimer's disease or Parkinson's disease). This is the amino acid
CC sequence of a metalloprotease MPl protein
XX
SQ Sequence 179 AA;

Query Match 25.8%; Score 549; DB 5; Length 179;
Best Local Similarity 66.5%; Pred. No. 3.8e-49;
Matches 119; Conservative 0; Mismatches 0; Indels 60; Gaps 1;

QY 38 IVLGIEISCDPTAAVDETVNVLGEAHSOTFVHLKTCGIYPPAAQQLHRENIQVGE 97
DB 1 IVLGIEISCDPTAAVDETVNVLGEAHSOTFVHLKTCGIYPPAAQQLHRENIQVGE 60
QY 98 ALSASGVSPDLSAATITIKPGIALSLGVLSFSLQVLQKKPFI----- 144
DB 61 ALSASGVSPDLSAATITIKPGIALSLGVLSFSLQVLQKKPFI----- 120
QY 145 -----IHMEAHALTIR 156
DB 121 GGAGGCTGATGCACTTACTATTAGGTGACCAATAAGATGATTTCTHHMEAHALTIR 179

RESULT 14
ABU22934
ID ABU22934 standard; protein; 382 AA.

AC ABU22934;

DT 19-JUN-2003 (first entry)

DE Protein encoded by prokaryotic essential gene #8461.

KW Antisense; prokaryotic essential gene; cell proliferation; drug design.

OS Bordetella pertussis.

PN WO20027183-A2.

PD 03-OCT-2002.

PF 21-MAR-2002; 2002WO-US009107.

PR 21-MAR-2001; 2001US-00815242.

PR 06-SEP-2001; 2001US-00948993.

PR 25-OCT-2001; 2001US-0342923P.

PR 08-FEB-2002; 2002US-00072851.

PR 06-MAR-2002; 2002US-0362699P.

XX (ELITRA PHARM INC.

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KU, Zyekind JW,

PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX MPI; 2003-029926/02.

DR N-PSDB; ACA26804.

XX Claim 25; SEQ ID NO 50858; 1766pp; English.

CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of

CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPD at
CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 382 AA;

Query Match 24.7%; Score 524; DB 6; Length 382;
Best Local Similarity 37.3%; Pred. No. 5.7e-46;
Matches 134; Conservative 63; Mismatches 136; Indels 26; Gaps 10;

QY 29 HPG--TLPLHK---IVLGIEISCDPTAAVDETVNVLGEAHSOTFVHLKTCGIYPPAA 83

DB 24 HPGPTLVNVSAPMTILGFSSCDPTGVAVCTERGLAHALMTQIAMHBYGVGVDELTA 83

QY 84 QQLHRENIQVGEALSASGVSPDLSAATITIKPGIALSLGVLSFSLQVLQKKPFI 143

DB 84 SRDHIRRVPLTRQVLAENGLTLADVGAVATVGPGLAGLTVGASVAAQLAWSBALPAI 143

QY 144 PIHMEAHALTIRLTN-KVEPPVLIIISGHCILALVGVSPFLIGKSLDPAQMD 202

DB 144 GIHHLGHLISPLLAEPREPFVALVSGHQMLMVDGVGYELLGETLIDPAAGAFD 203

QY 203 KVARRLSLIKHPECSTMSGKATEHLAKQGNRTHPDIKPELHAKNCDSPFTGLQ-HVTD 261

DB 204 KSAKLMGL-GYP-----GGPALARLAEQDASRYDLPRLHLSGDSDFSGIKTAVLT 256

QY 262 KITMKKEKEGIEKQIILSSAADIATVQHTMACHLVKTRTHAILFCQQRDLPPQNAVL 321

DB 257 RV--KAATRDGSELGR--QDRADLAATQAAYEVLAAKAIRL-----KQGL-----RL 304

QY 322 VASGVASNFYIRALEILTNAQTLLCPPLCTDNGIMIAMNGIERLPAGIGIHD 380

DB 305 VVAGVGAVANLIRAHILRALKPLRAEAYFPPLSLCTDNGAMIAFAAERYVAGIADURE 363

RESULT 15

ABG96491
ID ABG96491 standard; protein; 421 AA.

AC ABG96491;

DT 11-DEC-2002 (first entry)

DE Novel human metalloprotease associated protein #2.

CC Metalloprotease, MP-1; immune disorder; glutamate transport; cancer;
CC motor neuron disorder; amyotrophic lateral sclerosis; ALS; diabetes;
CC reproductive disorder; Klinefelter's syndrome; germinal cell aplasia;
CC genital wart; metabolic disorder; premature puberty; Kallman syndrome;
CC Cushing's syndrome; neurodegenerative disease; Alzheimer's disease;
CC Parkinson's disease; Huntington's disease; Tourette syndrome; sepsis;
CC liver disease; renal disease; immune disorder; rheumatoid arthritis;
CC acquired immunodeficiency syndrome; AIDS; pulmonary disease; pneumonia;
CC emphysema; cystic fibrosis; vascular disorder; inflammatory disorder;
CC neurological disorder.

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2005, 15:23:22 ; Search time 25.65 Seconds
(without alignment)
1552.972 Million cell updates/sec

Title: US-10-649-273-2

Perfect score: 2125

Sequence: 1 MLITTTAGVFFKPSKRRKY.....DISKEVGASIKVPOLKMEI 414

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_79:*

1: pirl1:*

2: pirl2:*

3: pirl3:*

4: pirl4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	634	29.8	463	2	B84888
2	553	26.0	387	2	E71711
3	548	25.8	365	2	AB2902
4	548	25.8	366	2	D97677
5	540.5	25.4	344	2	E97707
6	539.5	25.4	359	2	AB3274
7	535.5	25.2	367	2	PA7257
8	524	24.7	421	2	T18825
9	516.5	24.3	335	2	G70369
10	492.5	23.2	341	2	H83572
11	488.5	23.0	337	2	C91122
12	488.5	23.0	337	2	B85967
13	484.5	22.8	337	1	QOECR6
14	484.5	22.8	337	2	AG0892
15	483.5	22.8	337	2	A10079
16	475.5	22.4	325	2	A38108
17	475.5	22.4	342	2	H64074
18	469	22.1	340	2	B97011
19	467.5	22.0	354	2	C81040
20	465.5	21.9	346	2	F69786
21	465.5	21.9	354	2	C81986
22	463.5	21.8	343	2	D83718
23	460	21.6	336	2	C97888
24	459.5	21.6	326	2	H70195
25	457	21.5	327	2	G72411
26	457	21.5	326	2	E95015
27	455	21.4	348	2	S75548
28	451.5	21.2	344	2	AC1334
29	439	20.7	344	2	AB1705

30	439	20.7	346	2	AF1820	siatloglycoproteina
31	437.5	20.6	412	2	T40899	probable proteinase
32	431	20.3	348	2	D82807	O-sialoglycoprotei
33	430	20.2	341	2	G89996	hypothetical prote
34	409.5	19.3	338	2	A71545	probable O-sialogl
35	409	19.2	352	2	H71294	probable O-sialogl
36	406.5	19.1	344	2	H72106	O-sialoglycoprotei
37	406.5	19.1	344	2	B86515	O-sialoglycoprotei
38	405	18.8	374	2	T35581	probable O-sialogl
39	400.5	18.8	346	2	G86661	O-sialoglycoprotei
40	400	18.8	336	2	E84936	O-sialoglycoprotei
41	391.5	18.4	335	2	E81278	probable glycoprot
42	379.5	17.9	344	2	H70737	probable O-sialogl
43	361.5	17.0	302	2	S50740	OM17 protein - yea
44	361	17.0	302	2	F90526	hypothetical prote
45	353.5	16.6	315	2	A64205	O-sialoglycoprotei

ALIGNMENTS

RESULT 1

E84888 Probable O-sialoglycoprotein endopeptidase [imported] - Arabidopsis thaliana

C/Species: Arabidopsis thaliana (mouse-ear cress)

C/Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Jun-2003

C/Accession: E84888

R/Lin: X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Bentor, M.I.; Town, C.D.; Fujii, C.Y.;

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, J.

euss, D.; Niernan, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.

Nature 402, 761-768, 1999

A/Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A/Reference number: A84420; NCID:20083487; PMID:10617197

A/Accession: E84888

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-463 <STO>

A/Cross-references: GB:A8002093; NID:g2583127; PIDN:AAB82636.1; GSPDB:GN00139

C/Genetics:

A/Gene: A82945270

A/Map position: 2

C/Superfamily: O-sialoglycoprotein endopeptidase

Query Match	Score	Length	ID	Description
Best Local Similarity	37.9%	Pred. No. 2e-43:		
Matches 145; Conservative 61; Mismatches 129; Indels 48; Gaps 7;				
QY	38	IVLGIEISCDPTAAAVVDETVGAVLGEAHSQTEVHLKTVGIIVPAAGQOLHRENIQRIYQV	97	
DB	85	VVLGIEISCDPTAAAVVDETVGAVLGEAHSQTEVHLKTVGIIVPAAGQOLHRENIQRIYQV	141	
QY	98	ALSASGVSPSDLSAIVTTTKPGIALSLGGLSLQVGLQKPPPIHHHMEHALTIRL	157	
DB	142	ADDKNLTREKDSAVAVVTTPGSLCTLRGVRARVAVGSPSPPIVGHVHMEHALVAVL	201	
QY	158	T-NKVEPPLVLLISGHCCLALVQGVSDPLIGKSLDIAFGDMLKVARRSLIKHPEC	216	
DB	202	VQOELSPFMALLISGHCCLALVQGVSDPLIGKSLDIAFGDMLKVARRSLIKHPEC	258	
QY	217	STMSGKAIEHLAKQGNRHFPIKPLIHAQNCDFSTGLOHTYDKIMKEKEEGIEKG	276	
DB	259	--RSGPRAVELALBEDASVKNVPMKTHKCNFVAGIKQVRLAIEKE-----	308	
QY	277	QLTSSADIAATVQHTMACHLYKTRTRAILFCQRDLDPONNAVLAASGVASNPFIYRRA	336	
DB	309	--IRNRADIAASFQVAVVHLSEKCRRAIDWALE--LEPSIRGMVYISGVASNPFIYRRL	363	
QY	337	LEILTNATQCTLLCPPRFCTDNGIMIANNGIFRLAPAGLILHDIGIRVE-----	387	
DB	364	LNIVENKRLKIVCPSPSLCTDNGVAVVATGLHFRVG-----RVDPPPATPEP	413	
QY	388	-----PKCPGVSDISKVEGEA 403		


```

QY 155 ILLTNVPEPELVILISGHCILLANOGVSPELLGKSJLDIAKCDMLDKVARTSLIKP 214
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 128 ARTDGSFPEYMLVSGHQTLVLRGVGEIERMGTITDGLGEAFDTAKLGL--PYP 186
QY 215 ECSTWSGKAIEHLAKOGNRFHFDIKPELHAKNCDFSEFTG---QHVTDKIIMKEKE 270
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 187 -----GSPAVENAAKGDPRFPLPRMVEGARLDSPSGLKTAVRQATATAIPSEQD 240
QY 271 EGIEKQILSSADIATVQHTMACHLVKETHRALLFCQORULLPQNNNA--VIVASGVVA 328
Db 241 -----IADICASFQKAVSRTLLKORIGRLARFVE--PPIINSEPALVAVGVVA 287
QY 359 SNFYIRRAEILTNNATQCTLLCPPEPLCTDNGIMIAMNGIEERLRGILGIIHDEIGRIYEP 388
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 288 ANOEIRQTLQALCDTHGRFVFAVPHRLCTDNNAMIAMGLERMAEG---RQADALEVAP 343
QY 369 KCPLGVDSIKE 399
Db 344 KSRWPLDGSNE 354

```

RESULT 5

E97707
O-sialoglycoprotein endopeptidase (EC 3.4.24.57) [imported] - Rickettsia conorii (strain A:Rickettsia conorii)
C:Species: Rickettsia conorii
C>Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 22-Oct-2001
C:Accession: E97707
P:Orgn: H.; Audit: S.; Renesto-Audiffren, P.; Fournier, P.E.; Barde, V.; Sanson, D.; Rodiere, J.-M.; Audic, S.; Renault, C.; Baudouin, O.; Lippman, M.; Gaudin, P.; Lippman, M.; Science 293, 2098-2106, 2001
A>Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
A:Reference number: A97700; MUID:21442074; PMID:11557893
A:Accession: E97707
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-344 <KUR>
A:Cross-references: GB:AE006914; PIDN:AAL0259.1; PID:g15619097; GSPDB:GN00173
C:Genetics:
A:Gene: gcp
A:Superfamily: O-sialoglycoprotein endopeptidase
A:Keywords: hydrolase, metalloproteinase

Query Match	25.4%	Score 540.5	DB 2	length 344
Best Local Similarity	37.4%	Pred. No. 4.9e-36		
Matches 125	Conservative 58	Mismatches 132	Indels 19	Gaps 5

Qy	39	VIGIETSDCPAAAVVDEGVNIGVAHISQTEVHLKTGIIYPPAAQOLRENIQRTVCSA	98
		: : : : : : : : : :	
Db	4	ILGIBSSCDDBRAVSIITERRETILSNIIISQNTTEHAFGVGEIARSHLSHDKAKV	63
Qy'	99	LSASGVSPDSISAIATTTIKGIALISGVLSFSLQVGLKKPFIPIHHMEAHALTRLT	158
		: : : : : : : : :	
Db	64	LKESHTKLTIDISTIAATSGPGLIGVIVQSMARBSLKKPFIATNHLBGHALTRLT	123
Qy	159	NKVEPPLVLIIISGCHLALVQVSDFLLEKSLDIAPGMDLKVARRLSLIKAPECST	218
		: : : : : : : : :	
Db	124	DNIPPYLLILASGHCQFVAITGLGKKTLIGSTIDDAANGAFDKVAKMLNT-----A	176
Qy	219	MSGKAIIHLAKQGNRFHFDIKPRLHAHNCDSFTGLQHTYTDKIIMK-KEKEGIEKQ	277
		: : : : : : : :	
Db	177	PPGGEIEKRKALGDPHKYKPKPFIINSNCMSGSLKTAVRTLIMTLKEINDTV----	232
Qy	278	ILSSADIDIAVQHTMACHLVKTRALLFCQ-RDLLPQNNAVLVASGVASNFYIRRA	336
		: : : : : : : :	
Db	233	-----INDIAASQFPTIGELISSKQDARAYEQITNNPDKN--IYIAGVANNKYLQKI	286
Qy	337	LEILTNATQCTLGCPRLCTDNGIMIAMNGER	370
		: : : : : : : :	
Db	287	LSSCAKTYGRILYPIPHLCTDNAMIAVAGIER	320

RESULT 6

AB3274

O-sialoglycoprotein endopeptidase (EC 3.4.24.57) [imported] - Brucella melitensis (strain C)Species: Brucella melitensis
C.Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 15-Feb-2002
C.Accession: AB3274
R.DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova, E.; Mazur, M.; Goldsman, E.; Selkov, E.; Elizer, P.H.; Hagius, S.; O'Callaghan, D.; Leveson-Proctor, Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A.Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis
A.Reference number: AD3252; PMID:11756688
A.Accession: AB3274
A.Status: preliminary
A.Molecule type: DNA
A.Residues: 1359 <KUR>
A.Cross-references: GB:AA008917; PTDN:AA51357.1; PID:G17982056; GSPDB:GN00190
A.Experimental source: Strain 16M
C1Genetics:
A:Gene: BMEI0175
A:Map position: I
C:Superfamily: O-sialoglycoprotein endopeptidase
C:Keywords: hydrolase; metalloproteinase

Query Match	25.4%;	Score 539.5;	DB 2;	Length 359;
Best Local Similarity	38.6%;	Pred. No. 6.3e-36;		
Matches 132;	Conservative 52;	Mismatches 127;	Indels 31;	Gaps 7

```

OY 39 VLGITSCDDTFAAAVDE-----TGNVIGBAHISQTEVHLKTGGIVPPAAQOHLRENIQRI 94
Db 3 VLGIEITSCDBETAAALVEERDDMEGGRILSNVNLSQLAHEHPYGGVVEPIARAAVEALDR 62
OY 95 VOEALASAGVSPSDLSAIAITTIKPGTALSIGVLSFSLVGCQJLKPFIPIHHEMEAHALT 154
Db 63 VDRALNDAGLKYEDVAATAAGPGLIGSLVGLMTAKALMAAAQEPFAVANHBEHALT 1222
OY 155 IRLTNKVEFPLVLLISGGHCLLALVOGVSDFLLKSLDIAIPGMDLKVARRLSIKNP 214
Db 123 ARLTGFLPFPYLLLVSGHTQMWLVRGIDYERIGTTIDDLGAEFDTAKLGL-PPY 181
OY 215 ECSTMSGGKAEHLAKQGNRRHFPIKPRLHAAKNCDSFTGL-----QHVTDKIKKEX 265
Db 182 -----GGPAVERMALQGGQKRFALPRPLKGEARLDPSFSGKTAVRQATELVPLTDQ- 234
OY 270 EBGIEKGILSSADIAATVQHTMACHLVKRRHRAILPFCKQBDLPQ-NNAVLVASAGVA 328
Db 235 -----DVTDICASFQAAVAVDLTSDRAGRSIERPTE--FPDCATPSPVLVAGVA 281
OY 329 SNFYIRBALBITNATQCTLLCPBRLCTDNGIMIAWNCIER 370
Db 282 ANKTLRAALENICTRHGPAFLAPRLNLCTDNAMAMIAWAER 323

```

RESULT 7

F87257
 peptidase M22 family protein [imported] - Caulobacter crescentus
 C:Species: Caulobacter crescentus
 C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #ext_change 10-May-2001
 C:Accession: F87257
 B:McMern, W.C.; Feldbllym, T.V.; Paulsen, I.T.; Nelson, K.B.; Eisen, J.; Heid
 B.; Labb, M.C.; DeBo, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.
 n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Pr
 Proc. Natl. Acad. Sci. U.S.A. 98: 4136-4141, 2001
 A:Title: Complete Genome Sequence of *Caulobacter crescentus*.
 A:Reference number: AB7249, MUID:21173698, PMID:11259647
 A:Accession: F87257
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-367 <STO>
 A:Cross-references: GB:AA005673; NID:g13421168; PIDN:AAK22058.1; GSPDB:GN00148
 C:Genetics:
 A:Gene: CC0071
 A:Superfamily: O-sialoglycoprotein endopeptidase

Query Match	25.2%;	Score 535.5;	DB 2;	length 367;
Best Local Similarity	37.9%;	Pred. NO. 1.4e-35;		

Matches	136;	Conservative	54;	Mismatches	144;	Indels	25;	Gaps	7;
Qy	38	IVLGIEFSCDPTAAAV----	DEGNVLGEAIHISQTEVHLKTGGIVPPAAQOLHRENIOR	93					
Db	10	LIILGIEFSCDPTAAASV	RRADGVTVTLSVYIGQFEBHAPFGGVPEIARAHVESIDA	69					
Qy	94	IVQGLSASGVS	PSPLSAITTTIKPGTALSIVGVGSFSLQVQGLKKPPIPIHMEAHAL	153					
Db	70	IAAAVRAAGV	FGFDLDGVAATAGBGLVGVWGLAREKAVALLARGAPLVAVNHLBEGHAY	129					
Qy	154	TIRLTNKEFP	PLVLLISGGHCLTALVGVSDPFLLGKSLDIAEGDMLDKVARRLSLIKH	213					
Db	130	SARIGADIAV	PFLLLVGGHQQLLEVSGVCAKCRKLTGTTIDAGAEPADIKAKSLGL--PY	188					
Qy	214	PECSMTSGGKA	IEHLAKGNRFHEDIKPEPLHAAKCDPFSFGLQHTVDKTIIMKKEKEGI	273					
Db	189	P-----	GGPALKEKLAVGGDDPTRYALPRALLGRKDCDFSFSGLTAAARLAETLTDD--	240					
Qy	274	EKGQILSAA	IIAATVQHTMACHVYKFRHRAILPCQKRDLLPQNNAVLVASGVASNFYI	333					
Db	241	-----	ARRDLAAGVQAIARQULSERDRMKLYK--DSHDPEDLRFVAGVAANGAV	291					
Qy	334	RALEILTNA	TOCTLLCCPPRLCTDNGIMIAMNGIERLRAGILHDIIEGIRYEPKCP	392					
Db	292	RAALLADEK	KGFSFAAPPLAVCTDNAMIMLAGERU--ALGIFDDLDALA-RPRNPL	347					
RESULT 8									
T18825									
hypothetical protein C01G10.10 - Caenorhabditis elegans									
C:Species: Caenorhabditis elegans									
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Mar-2000									
C:Accession: T18825									
R:Matthews, L.									
submitted to the EMBL Data Library, October 1996									
A:Reference number: Z19027									
A:Accession: T18825									
A:Status: preliminary; translated from GB/EMBL/DBJ									
A:Molecule type: DNA									
A:Residues: 1-421 <EMBL>									
A:Cross-references: EMBL:Z81030; PIDN: CAB02716.1; GSPDB: GN00023; CESP: C01G10.10									
A:Experimental source: clone C01G10									
C:Genetics:									
A:Gene: CESP:C01G10.10									
A:Map position: 5									
A:introns: 31/2; 72/3; 122/2; 177/1; 272/3; 315/1; 353/2									
C:Superfamily: O-sialoglycoprotein endopeptidase									
Query Match									
		24.7%;	Score 524;	DB 2;	Length 421;				
		Best local similarity	33.4%;	Pred. No. 1.4e-34;					
		Matches 130;	Conservative 77;	Mismatches 146;	Indels 36;	Gaps 8;			
Qy	39	VLGIEFSCDPTAAAV	VDETGNVLGEAHSQTEVHLKTGGIVPPAAQOLHRENIOR	VOEA	98				
Db	25	VLGIEFSCDPTAAV	LVNKEKREILSERYTERALIQGGGINSVCALQHRENLRLIKC	84					
Qy	99	LSASGVSPSD	LSAIAATTIKPGTALSIVGVGSFSLQVQGLKKPPIPIHMEAHAL	TLRLT	158				
Db	85	LMDAGTSEK	DDAVAVATVPGVLVIAKEGISAIIGFAKKHRLPLIPVHMDAHAL	SLILV	144				
Qy	159	NKYEFPLV	LLISGGHCLTALVGVSDPFLLGKSLDIAEGDMDKVARRLSLIKHPES</						

RESULT 9

G70369
sialoglycoproteinase - Aquifex aeolicus
C:Species: Aquifex aeolicus
C:date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 09-Jul-2004
C:Accession: G70369
R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ov
V.
Nature 392, 353-358, 1998
A>Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID:9819666; PMID:9537320
A:Accession: G70369
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-335 <AO>
A:Cross-references: UNIPROT:O66986; GB:AEO00708; NID:g2983356; PIDN:AACO6951.1; PID:g298
A:Experimental source: strain VFS
C:Genetics:
A:Gene: gcp
C:Superfamily: O-sialoglycoprotein endopeptidase

Query Match 24.3%; Score 516.5; DB 2; Length 335;
Best Local Similarity 36.5%; Pred. No. 4.2e-34;
Matches 130; Conservative 66; Mismatches 125; Indels 35; Gaps 10;

Dy 40 LGIETSCDDTAADVDEGNVLGEAIHSQTEVHLKGGIVPPAAQQHRENIQRIVOEAL 99
 |::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 4 LAVENTCDETAIAIYDOKGVLCNVILSQAVHSPFGVVPDELARERTNTLPFDRL 63
 |::||::||::||::||::||::||::||::||::||::||::||::||
Dy 100 SASGVSPSDLASATIKRGIALSLGVGSFSLQVQLKKPFIPIHMEAHALTIRLTN 159
 |::||::||::||::||::||::||::||::||::||::||::||::||
Db 64 KESRNINLEIDISFLTPGLISTLVGVAFAKLAAYEKPLPVNHLEGHYSVFLEK 123
 |::||::||::||::||::||::||::||::||::||::||::||::||
Dy 160 KVEPFVLVLLSGHGCLALVQGVDFLLGKSLLDIAPGMLDKVARLSLKHPECSTM 219
 |::||::||::||::||::||::||::||::||::||::||::||::||
Db 124 KYEVFFLLIISSGHVDLYVRPGRYDPFGCTIDDVAAGEAYDKAKMLGL-GYP----- 177
 |::||::||::||::||::||::||::||::||::||::||::||::||
Dy 220 SGGKAIEHLAKQNRHFEDIKPPLHAANKCDFSTGLQHTVTDKIIMKKKEBGEIEKGQTL 279
 |::||::||::||::||::||::||::||::||::||::||::||::||
Db 178 -GGPIIDRLAKGKGL-YPLPKPELMERGINFPSSGLK---TALLNLKKEGNVRK---- 228
 |::||::||::||::||::||::||::||::||::||::||::||::||
Dy 280 SSAADIIAIVQHTMACHLVKTRTHALLFCRKORDLLPONNAVLSAGVASNYIRRALEI 339
 |::||::||::||::||::||::||::||::||::||::||::||::||
Db 229 --EDIAVSFOETVAILERS---LWAMKKTGIKR---LVVGVSANSRLR--EV 274
 |::||::||::||::||::||::||::||::||::||::||::||::||
Dy 340 LTNATO---CTLICPPRLCTDNGIMIAMWGIBELRAGILGILDIEGIRYEKCP 392
 |::||::||::||::||::||::||::||::||::||::||::||::||
Db 275 FKGAQGYEPFLYIPHPSLSTDNALMIAYAGMERFKGVAVPLDVNP---QNIFL 327
 |::||::||::||::||::||::||::||::||::||::||::||::||

RESULT 10

H83572
O-sialoglycoprotein endopeptidase PA0580 [imported] - Pseudomonas aeruginosa (strain PAO
C:Species: Pseudomonas aeruginosa
C:date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: H83572
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Bra
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kae, A.; Larbig, K.; Lim,
J.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A>Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: H83572
A>Status: preliminary
A:Molecule type: DNA

[illegible]

Db 230 --QTRADIARAFEDAVVDTLAIKSKRA-----LDOTGPKRLVIAGVGSANOTLRUK 278
Qy 337 LEILITNATQCTLLCPPRLCTDNGIMIAMNGIERLRAGILHDIEGIRYEPKCPU 392
Db 279 LADMMOKRGGEVFFYARPEFCTDNGAMIAVAGMVRRLRSN---LNSELSVSVRPRWPL 331

Search completed: November 10, 2005, 16:53:07
Job time : 26.65 secs

Protein Sequence Searches - February 2005

All of the sequence databases on ABSS have recently been updated.

- Please note that the curators of the UniProt database have purged some temporary accession numbers from the most recent version of UniProt. These sequences have been assigned new permanent accession numbers. The new UniProt record may not contain the previous temporary accession number.
- If you encounter an accession number from an older search run against UniProt (results file extension .rwp) that can no longer be found in the database, the permanent record with the new accession number can be found by searching the old accession number in the UniProt Protein Archive database (UniPARC) at:

<http://www.pir.uniprot.org/database/archive.shtml>

If you have any questions regarding this information or your results, please contact any STIC searcher.

When submitting sequence search results for scanning into IFW, please include a copy of this attachment to assist any future Examiners or members of the public who may encounter UniProt temporary accession numbers.

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2005, 15:16:01 ; Search time 100.35 Seconds

(without alignments)
2112.614 Million cell updates/sec

Title: US-10-649-273-2

Perfect score: 2125

Sequence: 1 MLILTTAGVFPKSKRKVY.....DISKEVGASIKVQLKMEI 414

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt_03.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2125	100.0	414	096EV9	096EV9 homo sapien
2	2090.5	98.4	439	09H4B0	09H4B0 homo sapien
3	1845	86.8	364	096NM5	096NM5 homo sapien
4	1835	86.4	414	096PB4	096PB4 mus musculu
5	1827	86.0	414	096LB6	096LB6 mus musculu
6	1819	85.6	414	09D0N0	09D0N0 mus musculu
7	1725	81.2	467	06AYN7	06AYN7 rattus norv
8	1283	60.4	404	08JFW3	08JFW3 brachydantio
9	1279	60.2	404	08JFR7	08JFR7 brachydantio
10	772.5	36.4	401	07QJ98	07QJ98 anopheles g
11	714.5	33.6	409	09VWD6	09VWD6 atropisophila
12	668.5	31.5	480	022145	022145 atropisophila
13	577.5	27.2	335	073H71	073H71 wolbachia p
14	568.5	26.8	360	092LH8	092LH8 rhizobium m
15	564	26.5	362	098ER6	098ER6 rhizobium l
16	558	26.3	387	068XK3	068XK3 rickettsia
17	553	26.0	387	068XK3	068XK3 rickettsia
18	548	25.8	365	08UC47	08UC47 agrobacteri
19	548	25.8	366	07CWJ8	07CWJ8 agrobacteri
20	541.5	25.5	344	07PAG7	07PAG7 rickettsia
21	540.5	25.4	344	092JX6	092JX6 rickettsia
22	539.5	25.4	359	08YJBI	08YJBI brucella me
23	539.5	25.4	359	08FYI5	08FYI5 brucella su
24	538.5	25.3	323	0960S6	0960S6 atropisophila
25	535.5	25.2	367	09ABZ9	09ABZ9 calobacter
26	531.5	25.0	389	06NDS4	06NDS4 rhodospseudo
27	524	24.7	421	093170	093170 caenothabdi
28	520	24.5	340	06FCR3	06FCR3 actinobact
29	516.5	24.3	335	07VXN4	07VXN4 aquifex aeo
30	515.5	24.3	364	07VXN4	07VXN4 bordetella
31	512	24.1	364	06GIR3	06GIR3 bartonella

32	510.5	24.0	346	2	07W668	07W668 bordetella
33	510	24.0	339	2	06LV10	06LV10 photobacter
34	509.5	24.0	346	2	07W134	07W134 bordetella
35	507.5	23.9	357	2	089WM1	089WM1 bradyrhizob
36	503.5	23.7	341	2	07NUE3	07NUE3 chromobacte
37	494.5	23.3	343	2	09CLJ1	09CLJ1 pasteurella
38	494	23.2	341	2	08ES16	08ES16 oceanobacil
39	492.5	23.2	337	1	GCP_SALTY	GCP_SALTY salmonella
40	492.5	23.2	341	2	0915V7	0915V7 pseudomonas
41	491	23.1	364	2	06FYF1	06FYF1 bartonella
42	490.5	23.1	337	2	08FDG6	08FDG6 escherichia
43	489	23.0	353	2	07WQD9	07WQD9 candidatus
44	488.5	23.0	337	2	08XK3	08XK3 escherichia
45	486.5	22.9	337	2	082XN2	082XN2 nitrosomona

ALIGNMENTS

RESULT 1	PRELIMINARY:	PRT:	414 AA.
096EV9			
AC	096EV9		
DT	01-DEC-2001 (TREMBLrel. 19, Created)		
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)		
DT	01-MAR-2004 (TREMBLrel. 26, Last annotation update)		
DE	OSGEP1 protein.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Carnivora; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Muscle;		
RX	MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;		
RA	Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,		
RA	Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schaller G.D.,		
RA	Altschul S.F., Zeeberg B., Burow K.H., Scheffer C.F., Bhat N.K.,		
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,		
RA	Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,		
RA	Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,		
RA	Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,		
RA	Raha S.S., Loguettano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,		
RA	Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,		
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,		
RA	Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,		
RA	Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,		
RA	Whitling M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,		
RA	Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,		
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,		
RA	Krzyszinski M.I., Skalska U., Smailus D.E., Scherch A., Schein J.E.,		
RA	Jones S.J., Marra M.A.,		
RT	"Generation and initial analysis of more than 15,000 full-length human		
RT	and mouse cDNA sequences."		
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Muscle;		
RA	Strausberg R.L.		
RL	Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; BC011904; AAH11904.1; -		
DR	MEROPS; M2.004; -		
DR	GO; GO:0008450; F:0-sialoglycoprotein endopeptidase activity; IEA.		
DR	GO; GO:0008270; F:zinc ion binding; IEA.		
DR	GO; GO:0005058; F:proteolysis and peptidolysis; IEA.		
DR	InterPro; IPR000905; Peptidase M22.		
DR	InterPro; IPR009180; Pept M22_Osialgl.		
DR	Pfam; PF00814; Peptidase_M22; 1.		
DR	PIRSF; PIRSF004537; Osialgl_optda; 1.		
DR	PRINTS; PR002367; Osialgl_optda; 1.		
DR	PRODOM; P0002367; Peptidase_M22; 1.		
DR	TIGRFAMs; TIGR00329; GCP; 1.		
DR	SEQUENCE 414 AA; 45122 MW; A536B333F5C6B8DD CRC64;		

Query Match 100.0%; Score 2125; DB 2; Length 414;
 Best Local Similarity 100.0%; Pred. No. 2,6e-155;
 Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLITKTAGVFPKSKRYEFLRSFNHPGTLFLHKTIVLGIEISCDTAAAVDETGNV 60
 DB 1 LGEALHSOTVEHLKTKGIVPPAAQOHLHRENIQRIVOEALASAGVSPDSLAIATTKPG 120
 QY 61 LGEALHSOTVEHLKTKGIVPPAAQOHLHRENIQRIVOEALASAGVSPDSLAIATTKPG 120
 DB 61 LGEALHSOTVEHLKTKGIVPPAAQOHLHRENIQRIVOEALASAGVSPDSLAIATTKPG 120
 QY 121 ALSLGVSLSFSLQVLGOLKKEPPIPIHMEAHATLRLTNKVEFPPLVLLISGHCILALV 180
 DB 121 ALSLGVSLSFSLQVLGOLKKEPPIPIHMEAHATLRLTNKVEFPPLVLLISGHCILALV 180
 QY 181 QGVSDPFLILGKSLDIAFGMDLVKVARLSLIKHPECSTMSGKALIEHLAKOGNRFHPDIK 240
 DB 181 QGVSDPFLILGKSLDIAFGMDLVKVARLSLIKHPECSTMSGKALIEHLAKOGNRFHPDIK 240
 QY 241 PPLHAKNCDFSFYGLQHTVDKTIIMKKEKEBGIIEKQILSSAADIAATVOHTMACHLYKR 300
 DB 241 PPLHAKNCDFSFYGLQHTVDKTIIMKKEKEBGIIEKQILSSAADIAATVOHTMACHLYKR 300
 QY 301 THRALFCCKORDLPONNAVIVASGVASNFYIRRALILTNATQCTLLCPPRLCTDNG 360
 DB 301 THRALFCCKORDLPONNAVIVASGVASNFYIRRALILTNATQCTLLCPPRLCTDNG 360
 QY 361 IMIANGIERLRAGILHDIIEGIRYEPKCPGLVDISKEVGEASIKVPOLKMEI 414
 DB 361 IMIANGIERLRAGILHDIIEGIRYEPKCPGLVDISKEVGEASIKVPOLKMEI 414

RESULT 2

Q9H4B0 PRELIMINARY; PRT; 439 AA.
 AC 09H4B0;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Putative sialoglycoprotein type 2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Uterus;
 RA Chen J.M., Fortunato M., Barrett A.J.;
 RL Submitted (Oct-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ295148; CAC14666.1; -
 DR MEROPS; M22.004; -
 DR Genew; HGNC:23075; OSGEPL1.
 DR GO: GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.
 DR GO: GO:0008223; F:peptidase activity; IEA.
 DR GO: GO:0008270; F:zinc ion binding; IEA.
 DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR009180; Pept_M22_Osialgl.
 DR InterPro; IPR009180; Pept_M22_Osialgl.
 DR Pfam; PF00814; Peptidase_M22; 1.
 DR PIRSF; PIRSF004537; Osialglcptde; 1.
 DR PRINTS; PR00787; OSIALOPTASE.
 DR ProDom; PD002367; Peptidase_M22; 1.
 DR TrGFams; TIGR00329; gcp; 1.
 KW TIGRFAse.
 SQ SEQUENCE 439 AA; 48040 MW; 44849372C784B41P CRC64;

Query Match 98.4%; Score 2090.5; DB 2; Length 439;
 Best Local Similarity 93.8%; Pred. No. 1.3e-152;
 Matches 412; Conservative 0; Mismatches 2; Indels 25; Gaps 1;
 QY 1 MLITKTAGVFPKSKRYEFLRSFNHPGTLFLHKTIVLGIEISCDTAAAVDETGNV 60

DB 1 MLITKTAGVFPKSKRYEFLRSFNHPGTLFLHKTIVLGIEISCDTAAAVDETGNV 60
 QY 61 LGEALHSOTVEHLKTKGIVPPAAQOHLHRENIQRIVOEALASAGVSPDSLAIATTKPG 120
 DB 61 LGEALHSOTVEHLKTKGIVPPAAQOHLHRENIQRIVOEALASAGVSPDSLAIATTKPG 120
 QY 121 ALSLGVSLSFSLQVLGOLKKEPPIPIHMEAHATLRLTNKVEFPPLVLLISGHCILALV 180
 DB 121 ALSLGVSLSFSLQVLGOLKKEPPIPIHMEAHATLRLTNKVEFPPLVLLISGHCILALV 180
 QY 181 QGVSDPFLILGKSLDIAFGMDLVKVARLSLIKHPECSTMSGKALIEHLAKOGNRFHPDIK 240
 DB 181 QGVSDPFLILGKSLDIAFGMDLVKVARLSLIKHPECSTMSGKALIEHLAKOGNRFHPDIK 240
 QY 241 PPLHAKNCDFSFYGLQHTVDKTIIMKKEKEBGIIEKQILSSAADIAATVOHTMACHLYKR 300
 DB 241 PPLHAKNCDFSFYGLQHTVDKTIIMKKEKEBGIIEKQILSSAADIAATVOHTMACHLYKR 300
 QY 301 THRALFCCKORDLPONNAVIVASGVASNFYIRRALILTNATQCTLLCPPRLCTDNG 360
 DB 301 THRALFCCKORDLPONNAVIVASGVASNFYIRRALILTNATQCTLLCPPRLCTDNG 360
 QY 361 IMIANGIERLRAGILHDIIEGIRYEPKCPGLVDISKEVGEASIKVPOLKMEI 414
 DB 361 IMIANGIERLRAGILHDIIEGIRYEPKCPGLVDISKEVGEASIKVPOLKMEI 414

RESULT 3

Q96NH5 PRELIMINARY; PRT; 364 AA.
 AC 096NH5;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Hypothetical protein FLJ30879.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
 RA Makamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
 RA Sekine M., Ohsashi M., Nishi T., Shibahara T., Tanaka T., Iehi S.,
 RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahari K.,
 RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,
 RA Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
 RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikawa E., Omura Y.,
 RA Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,
 RA Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,
 RA Tani H., Kimeta M., Watanabe M., Hirooka S., Chiba Y., Ishida S.,
 RA Ono Y., Takiguchi S., Watanabe S., Yosiida M., Hottu T., Kusano J.,
 RA Kanohori K., Takahashi-Fujii A., Hara R., Tanase T., Nomura Y.,
 RA Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,
 RA Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,
 RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,
 RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
 RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
 RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
 RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
 RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
 RA Ono T., Yamada K., Fujii Y., Ozaki K., Hiroo M., Omori Y.,
 RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
 RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
 RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
 RA Togaishi T., Oyama M., Hata H., Watanabe M., Komatsu T.,

RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuno Y., Yamashita R.,
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;
RT "Complete sequencing and characterization of 21,243 full-length human
RT cDNAs";
RL Nat. Genet. 36:40-45(2004).
DR EMBL: AK055441, BAB70923.1; -.
DR MEROPS: M22.004; -.
DR GO: GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.
DR GO: GO:0008233; F:peptidase activity; IEA.
DR GO: GO:0008270; F:zinc ion binding; IEA.
DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro: IPR000905; Peptidase_M22.
DR InterPro: IPR009180; Pept_M22_Osialgl.
DR Pfam: PF00814; Peptidase_M22; 1.
DR PIRSF: PIRSF004537; Osialglc_ptds; 1.
DR PRINTS: PR00789; OSIALOPTASE.
DR Prodom: PD002367; Peptidase_M22; 1.
DR TIGRFAMs: TIGR00329; gcp; 1.
KW Protease.
SQ SEQUENCE 364 AA; 39528 MW; E0E605A07D0EC3D6 CRC64;

Query Match 86.8%; Score 1845; DB 2; Length 364;
Best Local Similarity 99.2%; Pred. No. 8,4e-134;
Matches 361; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MLITKTAGVFPKSKRYEFLRSFNFPGTLFLHKIVLGIETSCDDTAAAVDEGNV 60
Db 1 MLITKTAGVFPKSKRYEFLRSFNFPGTLFLHKIVLGIETSCDDTAAAVDEGNV 60
QY 61 LGEAHSQTEVHLKTGGIVPPAAQQLHRENIQRIYQELASASGVSPDLSAATTIKPGL 120
Db 61 LGEAHSQTEVHLKTGGIVPPAAQQLHRENIQRIYQELASASGVSPDLSAATTIKPGL 120
QY 121 ALSIGVLSFSLQVQLKKPFIPIHMEAHALTIRLTNKVPPFVLVILSGHCLLALV 180
Db 121 ALSIGVLSFSLQVQLKKPFIPIHMEAHALTIRLTNKVPPFVLVILSGHCLLALV 180
QY 181 QGVSDFLLGKSLDIAPGMDLVKARSLSLIKHPECSMTSGSKAIEHLAKQGNRFHFDIK 240
Db 181 QGVSDFLLGKSLDIAPGMDLVKARSLSLIKHPECSMTSGSKAIEHLAKQGNRFHFDIK 240
QY 241 PPLHAKNCDFSFPTGLQHTVDKIIMKKEKEGIEKQIILSSADIAATVQHTMACHLYKR 300
Db 241 PPLHAKNCDFSFPTGLQHTVDKIIMKKEKEGIEKQIILSSADIAATVQHTMACHLYKR 300
QY 301 THRALLFCQKRDLLPNNNAVIVASGVASNFYIRALFILTNAQTCTLLCPPRLCTDNG 360
Db 301 THRALLFCQKRDLLPNNNAVIVASGVASNFYIRALFILTNAQTCTLLCPPRLCTDNG 360
QY 361 IMIA 364
Db 361 IMIA 364

RESULT 4

Q6PEB4 PRELIMINARY; PRT; 414 AA.
AC Q6PEB4;
DT 05-JUL-2004 (TREMBLrel. 27, Created)
DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C26G II; TISSUE=Mammary tumor;
RX MEDLINE=2238257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang Y., Hsieh F.,
RA Diachenko L., Mausina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Schaefer C.,
RA Raha S.S., Locantello N.A., Peters G.J., Abramson R.D., Mullaney S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.U., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.W., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalins D.E., Scherch A., Schein J.E.,
RA Jones S.J., Maitra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C26G II; TISSUE=Mammary tumor;
RA Strausberg R.;
RL Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC058172; AAH58172.1; -.
DR GO: GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.
DR GO: GO:0008270; F:zinc ion binding; IEA.
DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro: IPR000905; Peptidase_M22.
DR InterPro: IPR009180; Pept_M22_Osialgl.
DR Pfam: PF00814; Peptidase_M22; 1.
DR PIRSF: PIRSF004537; Osialglc_ptds; 1.
DR PRINTS: PR00789; OSIALOPTASE.
DR Prodom: PD002367; Peptidase_M22; 1.
DR TIGRFAMs: TIGR00329; gcp; 1.
KW Hypothetical protein.
SQ SEQUENCE 414 AA; 44962 MW; 0418AD93C6FEBCAE CRC64;

Query Match 86.4%; Score 1835; DB 2; Length 414;
Best Local Similarity 85.0%; Pred. No. 5,8e-133;
Matches 352; Conservative 25; Mismatches 37; Indels 0; Gaps 0;

QY 1 MLITKTAGVFPKSKRYEFLRSFNFPGTLFLHKIVLGIETSCDDTAAAVDEGNV 60
Db 1 MLITKTAGVFPKSKRYEFLRSFNFPGTLFLHKIVLGIETSCDDTAAAVDEGNV 60
QY 61 LGEAHSQTEVHLKTGGIVPPAAQQLHRENIQRIYQELASASGVSPDLSAATTIKPGL 120
Db 61 LGEAHSQTEVHLKTGGIVPPAAQQLHRENIQRIYQELASASGVSPDLSAATTIKPGL 120
QY 121 ALSIGVLSFSLQVQLKKPFIPIHMEAHALTIRLTNKVPPFVLVILSGHCLLALV 180
Db 121 ALSIGVLSFSLQVQLKKPFIPIHMEAHALTIRLTNKVPPFVLVILSGHCLLALV 180
QY 181 QGVSDFLLGKSLDIAPGMDLVKARSLSLIKHPECSMTSGSKAIEHLAKQGNRFHFDIK 240
Db 181 QGVSDFLLGKSLDIAPGMDLVKARSLSLIKHPECSMTSGSKAIEHLAKQGNRFHFDIK 240
QY 241 PPLHAKNCDFSFPTGLQHTVDKIIMKKEKEGIEKQIILSSADIAATVQHTMACHLYKR 300
Db 241 PPLHAKNCDFSFPTGLQHTVDKIIMKKEKEGIEKQIILSSADIAATVQHTMACHLYKR 300
QY 301 THRALLFCQKRDLLPNNNAVIVASGVASNFYIRALFILTNAQTCTLLCPPRLCTDNG 360
Db 301 THRALLFCQKRDLLPNNNAVIVASGVASNFYIRALFILTNAQTCTLLCPPRLCTDNG 360
QY 361 IMIANGIERLRAGGILHIDIGIRYEPKCPGVDSKEVGSASIVPOLKKEI 414
Db 361 IMIANGIERLRAGGILHIDIGIRYEPKCPGVDSKEVGSASIVPOLKKEI 414

RESULT 5

Q8BLB6 PRELIMINARY; PRT; 414 AA.
ID Q8BLB6

AC Q8BLB6; 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Mus musculus adult male corpora quadrigemina cDNA, RIKEN full-length
 DE enriched library, clone: B20219017 product: similar to PUTATIVE
 DE SIALOGLYCOPROTEINASE TYPE 2.
 GN Name=Osgp11;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;
 RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
 RA Carninci P., Hayashizaki Y.;
 RT "High-efficiency full-length cDNA cloning.";
 RL Meth. Enzymol. 303:19-44(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;
 RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
 RA RIKEN PANTOM Consortium;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;
 RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
 RA The PANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs.";
 RL Nature 420:563-573(2002).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;
 RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
 RA Kono H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
 RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watabiki M.,
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system: 384-format
 RT sequencing pipeline with 384 multicapillary sequencer.";
 RL Genome Res. 10:1757-1771(2000).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;
 RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
 RA Kono H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
 RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watabiki M.,
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system: 384-format
 RT sequencing pipeline with 384 multicapillary sequencer.";
 RL Genome Res. 10:1757-1771(2000).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;
 RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
 RA Fukuda S., Furuno M., Hangaki T., Hara A., Hashizume W.,
 RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
 RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
 RA Katoh H., Kawai J., Kojima Y., Kondo S., Kono H., Konda M., Koya S.,
 RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
 RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
 RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
 RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
 RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
 RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK045669; BAC32450.1; -.

DR MEROPS; M22.004; -;
 DR MGD; MG1.1919335; Osgp11.
 DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.
 DR GO; GO:0008233; F:peptidase activity; IEA.
 DR GO; GO:0008270; F:zinc ion binding; IEA.
 DR GO; GO:0006508; F:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000905; Peptidase_M22.
 DR InterPro; IPR009180; Peptidase_M22_Osialgl.
 DR Pfam; PF00814; Peptidase_M22; 1.
 DR PIRSF; PIRSF004537; Osialglc_ptide; 1.
 DR PRINTS; PR00789; OSIALOPRASE.
 DR ProDom; PD002367; Peptidase_M22; 1.
 DR TIGRFAMs; TIGR00329; gcp; 1.
 KW Peptidase.
 SQ SEQUENCE 414 AA; 44927 MW; 9A2D27B576C7DB24 CRC64;
 Query Match 86.0%; Score 1827; DB 2; Length 414;
 Best Local Similarity 85.0%; Pred. No. 2.4e-132;
 Matches 352; Conservative 23; Mismatches 39; Indels 0; Gaps 0;
 QY 1 MLILRTAGVFFPKPSKRYEFLRSFNPFGTLFLHKIVLGJETSCDDTAAAVDETVGNV 60
 DB 1 MMLMRTAGAIRPKPSKRYGFLRRFSVHPRLISCHLVGLETSCDDTGAAYVDETGNV 60
 QY 61 LGEAHSQTEVALKTGGIVPPAAQQLARENIGRIVQELASAGSVSPSLSAIATTTKPGI 120
 DB 61 LGEALHSQTEVALKTGGIVPPAAQQLARENIGRIVETLSACRITPSDLSAATTTKPGI 120
 QY 121 ALSLVGVSFSGLVGLKRPPIPIHMEAAHLLTLTKVPEPFLVLLISGCHLALV 180
 DB 121 ALSLVGVSFSGLVGLKRPPIPIHMEAAHLLTLTKVPEPFLVLLISGCHLALV 180
 QY 181 QGVSPFLIGKSLDIAPGMDLKVARRSLIHPPECTSGSGGAIHLAKOGRFPFDIK 240
 DB 181 QGVSPFLIGKSLDIAPGMDLKVARRSLIHPPECTSGSGGAIQLADGRRFPFTIN 240
 QY 241 PELHAKNCDFFSTGLQHTVTDKTIIMKEKEBEGIKQILSSADIATVQHTWACHLVKR 300
 DB 241 PMQNAKNCDFFSTGLQHTVTDKTIIMKEKEBEGIKQILSSADIATAVQHTWACHLVKR 300
 QY 301 THRALLFCQKQMLSPANALVAVSGVANSFYIRALLETLTNAQTCTLLCPPLCTDNG 360
 DB 301 THRALLFCQKQMLSPANALVAVSGVANSFYIRKALEIVANVTCTLLCPPLCTDNG 360
 QY 361 IMIANGIERLRAGGIIHDIGIRPEKCPGVDISKVEGASIVPOLKMEI 414
 DB 361 IMIANGIERLRAGGIVLHVEDIRPEKCPGVDISKVEGASIVPOLKMEI 414
 RESULT 6
 ID Q9DONO PRELIMINARY; PRT; 414 AA.
 AC Q9DONO;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Mus musculus 10 days embryo whole body cDNA, RIKEN full-length
 DE enriched library, clone:2610001M19 product: similar to PUTATIVE
 DE SIALOGLYCOPROTEINASE TYPE 2.
 GN Name=Osgp11;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Whole body;
 RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
 RA Carninci P., Hayashizaki Y.;
 RT "High-efficiency full-length cDNA cloning.";
 RL Meth. Enzymol. 303:19-44(1999).
 RN [2]
 RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Whole body;
 RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
 RA RIKEN FANTOM Consortium;
 RT "Functional annotation of a full-length mouse cDNA collection."
 RL Nature 409:685-690(2001).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Whole body;
 RA the FANTOM Consortium,
 RT the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs."
 RL Nature 420:563-573(2002).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Whole body;
 RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/9r.145100;
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
 RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.,
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to
 RT prepare full-length cDNA libraries for rapid discovery of new genes."
 RL Genome Res. 10:1617-1630(2000).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Whole body;
 RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/9r.152600;
 RA Shibata K., Itoh M., Alzawa K., Nagaoka S., Sasaki N., Carninci P.,
 RA Kono H., Akiyama Y., Nishi K., Kitsuwa T., Tashiro H., Itoh M.,
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishino T., Harada A.,
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
 RA Fujiwara K., Inoue K., Ozawa Y., Izawa M., Ohara E., Watanabe M.,
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.,
 RT "RIKEN integrated sequence analysis (RISA) system-384-format
 RT sequencing pipeline with 384 multicapillary sequencer."
 RL Genome Res. 10:1757-1771(2000).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Whole body;
 RA Adachi T., Alzawa K., Akahira S., Akimura T., Arai A., Aono H.,
 RA Atakachi T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
 RA Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hirooka T., Hori F.,
 RA Imotani K., Ishii Y., Itoh M., Izawa M., Kaikawa T., Kato H.,
 RA Kawai J., Kojima Y., Kono H., Kouda M., Koya S., Kurahara C.,
 RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ono M.,
 RA Okazaki Y., Okido T., Owa C., Saito H., Saito R., Saki C., Saki K.,
 RA Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,
 RA Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
 RA Tejima Y., Toyota T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
 RA Muramatsu M., Hayashizaki Y.,
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 RL EMBL: AK011265; BAB27506.1; --
 DR MEROPS; M22.004; --
 DR MGD; MGI:191333; Osgp11.
 DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.
 DR GO; GO:0008233; F:peptidase activity; IEA.
 DR GO; GO:0008270; F:zinc ion binding; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR00905; Peptidase_M22.
 DR InterPro; IPR009180; Pept_M22_Osialgl.
 DR Pfam; PF00814; Peptidase_M22; 1.
 DR PIRSF; PIRSF004537; Osialglc_ptids; 1.
 DR PRINTS; PR00789; OSIALOPTASE.
 DR ProDom; PD002367; Peptidase_M22; 1.
 DR TIGRFAMs; TIGR00329; gcp; 1.
 DR TIGRFAMs; TIGR00329; gcp; 1.
 KM PROTEASE.
 SO SEQUENCE 414 AA; 44999 MW; 999BC689944DDB24 CRC64;

Query Match 85.6%; Score 1819; DB 2; Length 414;
 Best Local Similarity 54.0%; Freq. No. 5.3e-13;
 Matches 351; Conservative 23; Mismatches 40; Indels 0; Gaps 0;
 1 MLITKAGVGFKSKKRVFPLSFNHPGLFLHKIVLGIETSCDDTAAAVDEGNV 60

Db 1 MMLRRTAGIIPPEPKSKYGFRRFSVHRTLSCHLVIGITSCDDTAAAVDEGNV 60
 61 LGEAHSQTEVHLKGTGIIVPAQOLHRENIQRIVOEALASASGSPSDSAITTKPG 120
 61 LGEAHSQTEVHLKGTGIIVPAQOLHRENIQRIVOEALASASGSPSDSAITTKPG 120
 61 LGEAHSQTEVHLKGTGIIVPAQOLHRENIQRIVOEALASASGSPSDSAITTKPG 120
 121 ALSLVGLSFLSLQVLQKPKPIIHMEAAHTIRLTNRVPEPVLIISSGCHLALV 180
 121 ALSLVGLSFLSLQVLQKPKPIIHMEAAHTIRLTNRVPEPVLIISSGCHLALV 180
 121 ALSLVGLSFLSLQVLQKPKPIIHMEAAHTIRLTNRVPEPVLIISSGCHLALV 180
 181 QGVSDPDLILKSLIDIPAGMDLVNARLSIKKEPSTSGGALIEHLAQGRFFPD 240
 181 QGVSDPDLILKSLIDIPAGMDLVNARLSIKKEPSTSGGALIEHLAQGRFFPD 240
 181 QGVSDPDLILKSLIDIPAGMDLVNARLSIKKEPSTSGGALIEHLAQGRFFPD 240
 241 PPLHAKNCDFSPFGQHTYDKITMKKEEGEIKGQIISADIAVQHTMACHLVK 300
 241 PPLHAKNCDFSPFGQHTYDKITMKKEEGEIKGQIISADIAVQHTMACHLVK 300
 241 PPLHAKNCDFSPFGQHTYDKITMKKEEGEIKGQIISADIAVQHTMACHLVK 300
 301 THRAILFCQKQNLSPANVLVVGVSANLYIRKALIVANATQCTLLCPPRLCTD 360
 301 THRAILFCQKQNLSPANVLVVGVSANLYIRKALIVANATQCTLLCPPRLCTD 360
 301 THRAILFCQKQNLSPANVLVVGVSANLYIRKALIVANATQCTLLCPPRLCTD 360
 361 IMIANGIRLRAGLGLHDIGIRYEPKCPGLGVDISKEVGEASIVPOLKMEI 414
 361 IMIANGIRLRAGLGLHDIGIRYEPKCPGLGVDISKEVGEASIVPOLKMEI 414
 361 IMIANGIRLRAGLGLHDIGIRYEPKCPGLGVDISKEVGEASIVPOLKMEI 414

Db 361 IMIANGIRLRAGLGLHDIGIRYEPKCPGLGVDISKEVGEASIVPOLKMEI 414
 361 IMIANGIRLRAGLGLHDIGIRYEPKCPGLGVDISKEVGEASIVPOLKMEI 414
 361 IMIANGIRLRAGLGLHDIGIRYEPKCPGLGVDISKEVGEASIVPOLKMEI 414

RESULT 7
 ID 06AVN7 PRELIMINARY; PRT; 467 AA.
 AC 06AVN7;
 DT 25-OCT-2004 (TrEMBLrel. 28, Created)
 DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
 DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
 DE Hypothetical protein (Fragment).
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RX PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins P.S., Wagner L., Shennan C.M., Schlier G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heish F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Rana S.S., Loughran N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kerteman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green B.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Kravinsky M.I., Skalski U., Smalls D.E., Scherch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RA Director KGC Project;
 RL Submitted (Aug-2004) to the EMBL/GenBank/JNRI databases.
 DR EMBL; BC078974; AAH78974.1; --
 DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.
 DR GO; GO:0008270; F:zinc ion binding; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.

DR InterPro; IPR000905; Peptidase_M22.
 DR InterPro; IPR009180; Pept_M22_Osialgl.
 DR InterPro; IPR002016; Peptidase.
 DR Pfam; PF00814; Peptidase_M22; 1.
 DR PIRSF; PIRSF004537; Osialglc_pptds; 1.
 DR PRINTS; PR00789; OSIALOPTASB.
 DR ProDom; PD002367; Peptidase_M22; 1.
 DR TIGRFAMs; TIGR00329; gcp; 1.
 DR PROSITE; PS00435; PEROXIDASE_1; UNKNOWN_1.
 KW Hypothetical protein.
 FT NON_TER 467
 SQ SEQUENCE 467 AA; 50799 MW; 474E18B1959B8AC0 CRC64;

Query Match 81.2%; Score 1725; DB 2; Length 467;
 Best Local Similarity 86.1%; Pred. No. 2e-124;
 Matches 335; Conservative 19; Mismatches 35; Indels 0; Gaps 0;

QY 1 MLILTKTAGVFKKSKRYEFLRSFNHPGTLFLAKIVLGIEISCDDTAAAVVDETGVN 60
 DB 1 MLMSTKTAGAIPRPSPNVGRFIRFNVQPPALFHHKLVIGIEISCDDTAAAVVDETGVN 60
 QY 61 LGEAIIHSQTEVHLKTGGIVPPAQQIHRNTQRIYQELASGVSPPDLSAIAITIKKGL 120
 DB 61 LGEAIIHSQTEVHLKTGGIVPPAQQIHRNTQRIYQELASGVSPPDLSAIAITIKKGL 120
 QY 121 ALSTGVGVSFSLQVGO;KKRPPIPHHMEAHALTITLNNKVFPPVLVLLISGHCILALV 180
 DB 121 ALSTGVGVSFSLQVGO;KKRPPIPHHMEAHALTITLNNKVFPPVLVLLISGHCILALV 180
 QY 181 QGVSDPFLIGKSLDIPAGDMLDKVARRLSLIKHPECSMTSGGKALEHLAKQGNRPHFDIK 240
 DB 181 QGVSDPFLIGKSLDIPAGDMLDKVARRLSLIKHPECSMTSGGKALEHLAKQGNRPHFDIK 240
 QY 241 PPLHAAKNCDSFTGLQHVTDKIIIMKEKEBEGIEKGQILSSAADIATVQHTMACHLYKR 300
 DB 241 PPMONAKKCDPSFTGLQHVTDKILTHKEKEBEGIEKGQILSSAADIATVQHTMACHLYKR 300
 QY 301 THRATLFPKQKRDLLPQNNANVLVASGVASNFYIRALBELTMAQCTLLCPPRCTDNG 360
 DB 301 THRATLFPQOKLLSPANAVLVSGVASNLYIRALBELTMAQCTLLCPPRCTDNG 360
 QY 361 IMIANGIERLRAGIGLHIDIEGIRYEPK 389
 DB 361 IMIANGIERLRAGIGLHIDVEDIRYEPK 389

RESULT 8

Q8JFW3 PRELIMINARY; PRT; 404 AA.

AC O8JFW3; 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE SI:d2211013.4 (Novel glycoprotease).
 GN Name=d272B14.6;
 OS Brachydanio rerio (Zebrafish) (Danio rerio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Danio.
 OX NCBI_TaxID=7955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Babbase A.;
 RL Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL591593; CAD4347.1; -.
 DR MEROPS; M22.004; -.
 DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.
 DR GO; GO:0008233; F:peptidase activity; IEA.
 DR GO; GO:0008270; F:zinc ion binding; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR00905; Peptidase_M22.
 DR InterPro; IPR009180; Pept_M22_Osialgl.
 DR Pfam; PF00814; Peptidase_M22; 1.

DR PIRSF; PIRSF004537; Osialglc_pptds; 1.
 DR PRINTS; PR00789; OSIALOPTASB.
 DR ProDom; PD002367; Peptidase_M22; 1.
 DR TIGRFAMs; TIGR00329; gcp; 1.
 KW Protease.
 SQ SEQUENCE 404 AA; 43956 MW; 3A6B1DD050737B35 CRC64;

Query Match 60.4%; Score 1283; DB 2; Length 404;
 Best Local Similarity 64.4%; Pred. No. 1.9e-90;
 Matches 242; Conservative 58; Mismatches 76; Indels 0; Gaps 0;

QY 37 KIVLGIEISCDDTAAAVVDETGVNIGEAIIHSQTEVHLKTGGIVPPAQQIHRNTQRIYQ 96
 DB 25 RLVLGIEISCDTEGAVALDETERIIGESIIHSQTEHLKTGGIIPVLAQRIHRNTSRVQ 84
 QY 97 EALSASGVSPPDLSAIAITIKKGLALSTGVGVSFSLQVGO;KKRPPIPHHMEAHALTITR 156
 DB 85 EALNSALEPSELTVAVTKKGLALSTGLDYSILKTFVRQKQKPIPHHMEAHALTITR 144
 QY 157 LTNKVFPPVLVLLISGHCILALVQVSDPFLIGKSLDIPAGDMLDKVARRLSLIKHPEC 216
 DB 145 MHPIDFPFLVLLVSGHSLALAKGIDFLLGTLDBAAGTLDKIRRLSLRNHPEC 204
 QY 217 STMSGKALEHLAKQGNRPHFDIKQPLHAAKNCDSFTGLQHVTDKIIIMKEKEBEGIEKG 276
 DB 205 GTLSCGQAIERLAKEGDRLAFHFISPMQNYDCNFSFAGLRQITGAINKKEKEBEGVAG 264
 QY 277 QILSSAADIATVQHTMACHLYKRTTRATLFPKQKRDLLPQNNANVLVASGVASNFYIRRA 336
 DB 265 QPLSCVCKDIAAASQHTVASHLAKRTRATLFPCKSKGLPEQNTLLVSGVASNEFYIRQ 324
 QY 337 LEILTNATQCTLLCPPRCTDNGIMIANNGIERLRAGIGLHIDIEGIRYEPKPLGVDI 396
 DB 325 LKTIIDATLHLCPPSKCTDNGWMIANGIERLRKQKGLISYSSEVVEPPGAPLGDI 384
 QY 397 SKEVGEASIKVPOLKX 412
 DB 385 TSEVKEAAIKVPKRLX 400

RESULT 9

Q8JFR7 PRELIMINARY; PRT; 404 AA.

AC O8JFR7; 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE SI:d272B14.6 (Novel glycoprotease).
 GN Name=SI:d272B14.6;
 OS Brachydanio rerio (Zebrafish) (Danio rerio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Danio.
 OX NCBI_TaxID=7955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Clark G.;
 RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL672217; CAD43443.1; -.
 DR MEROPS; M22.004; -.
 DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.
 DR GO; GO:0008233; F:peptidase activity; IEA.
 DR GO; GO:0008270; F:zinc ion binding; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR00905; Peptidase_M22.
 DR InterPro; IPR009180; Pept_M22_Osialgl.
 DR Pfam; PF00814; Peptidase_M22; 1.
 DR PIRSF; PIRSF004537; Osialglc_pptds; 1.
 DR PRINTS; PR00789; OSIALOPTASB.
 DR ProDom; PD002367; Peptidase_M22; 1.
 DR TIGRFAMs; TIGR00329; gcp; 1.
 DR Protease.
 SQ SEQUENCE 404 AA; 44027 MW; 6FB98653A651860F CRC64;

Query Match 60.2%; Score 1279; DB 2; Length 404;
Best Local Similarity 64.1%; Pred. No. 3.9e-90;
Matches 241; Conservative 59; Mismatches 76; Indels 0; Gaps 0;

QY 37 KIVLGIENSCDDTAAAVVDEGTGNVGEAIIHSQTEVHLKTGGIVPPAAQOIHRENIQRIYQ 96
DB 25 RVLGIETSCDGTGAALVGTNGVLIHSQSSHLRFGGIIPVADIDIRANIESVYQN 84

QY 97 EALSASGVSPDSALIAITTTIKPGALSLGVGLSFLQVLGVKKPFIPIHMEAHALTYR 156
DB 85 EALNRSALIEBELAVATTVFGLALSLGIGLDYSLKFKRQKQKFIPIHMEAHALTYR 144

QY 157 LTNKVEPFPVLVLSGGHCLLALVGVSDPFLILGSLDIPAGDMLDKVARRLSLKHPEC 216
DB 145 MLHLDPFPVLVLSGGHSLALAKGIDBFLLGQTLDEAAGDTLKDIAARRLSLRNHPBC 204

QY 217 STWSGKRAIEHLAQGNRPHFDIKRPLHAKNCDSPFGLQVHTDKIMKKEKEGIEG 276
DB 205 GTLSGQAIEHLAKEGDQIAHFISPMQONYDCNFSFGLRQITGALNKKEKEGIEG 264

QY 277 QILSSADIAATVQHTMACHLVKRTHRAILFCRDLPPONNAVVASGVASNFYIRRA 336
DB 265 QFLSCVMDIAAASQHTVASHLAKRTHRAILFCRSGKLAPQONPTLIVSGVASNEYIRQI 324

QY 337 LEILTNATQCTLLCPPRLLCTDNGIMAMNGIERLRAGILHDIIEGIRYEPKCPGLVDI 396
DB 325 LKITDGTGHLTLCPSPKCTDNGVMAMNGIERLKOGGILSYSESVYEPKAPLGLDI 384

QY 397 SKEVGEASIKVPOLKM 412
DB 385 TSEVGEAAIKVPLKL 400

RESULT 10
QY 070918 PRELIMINARY; PRT; 401 AA.
ID 070918
AC 070918;
DT 01-MAR-2004 (Tremblrel. 26, Created)
DT 01-MAR-2004 (Tremblrel. 26, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE AGCP14990 (Fragment).
GN Name=agCG46164; ORFNames=ENSANG0000007922;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.
OX NCBI_TaxId=180454;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PEST;
RA Anopheles Genome Sequencing Consortium;
RU Submitted (MAR-2002) to the EMBL/Genbank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/Genbank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAAB01008900; EAA09387.1; -!
DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009095; Peptidase M22.
DR InterPro; IPR009180; Pept M22 Osa1gl.
DR Pfam; PF00814; Peptidase M22; 1.
DR PIRSF; PIRSF004537; Osa1glc; 1.
DR PRINTS; PR00789; OSA1LOPTASE.
DR ProDom; PD002367; Peptidase M22; 1.
DR TIGRFAMs; TIGR00329; gcp; 1.
FT NON TER 1
SQ SEQUENCE 401 AA; 43849 MW; CC9426723D3PD4FI CRC64;

Query Match 35.4%; Score 772.5; DB 2; Length 401;
Best Local Similarity 42.7%; Pred. No. 4.2e-51;
Matches 165; Conservative 68; Mismatches 142; Indels 11; Gaps 4;

QY 38 IVLGIETSCDDTAAAVVDEGTGNVGEAIIHSQTEVHLKTGGIVPPAAQOIHRENIQRIYQ 97
DB 9 VVLGIETSCDDTGAALVGTNGVLIHSQSSHLRFGGIIPVADIDIRANIESVYQN 68

QY 98 EALSASGVSPDSALIAITTTIKPGALSLGVGLSFLQVLGVKKPFIPIHMEAHALTYR 157
DB 69 AFLTAMPTPDIDAVAVTNRPGLPSLVGMRYAKHIANSYNRPILPIHMOAHALMARM 128

QY 158 TNKVEPFPVLVLSGGHCLLALVGVSDPFLILGSLDIPAGDMLDKVARRLSLKHPEC 217
DB 129 TSTIIPFPLCLVSGSHSLVFPESRARFLLGETDIDPAAGELDKIARRLKNVAKYA 188

QY 218 TWSGKRAIEHLAQG-----NRFHDIKRPLHAKNCDSPFGLQVHTDKIMKKEKEG 272
DB 189 QMSGQAIEHLAQGAKAKDTSAVEFPL--PLSKYRDCQSFAGLKNATATHILREESTLH 246

QY 273 IEKGQILSSADIAATVQHTMACHLVKRTHRAILFCRDLPPONNAVVASGVASNFYIRRA 330
DB 247 LAPDALLPVEAFCAFLKGVTRHMLRTORALEYCERRKLFSDAEHRSLVSGVACN 306

QY 331 FYIRRALIEILTNATQCTLLCPPRLLCTDNGIMAMNGIERLRRA--GIGILHDIIEGIRYEP 388
DB 307 DVIFNALSSMAAQFGSYSTRPPKGLCTDNGTMAMNGIELKADKOTAEWTTKTEQVDISG 366

QY 389 KCPLGVDISKEVGEASIKVPOLKMEI 414
DB 367 KCPIGDSLIDVGEANIAKMAKVDI 392

RESULT 11
QY 09VMD6 PRELIMINARY; PRT; 409 AA.
ID 09VMD6
AC 09VMD6;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE CG14231-PA.
GN ORFNames=CG14231;
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter B.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abriil J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktarcoglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borova D., Botchan M.R., Bouck J., Brodeur P., Brothier P.,
RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo S., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doull L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferrara C., Ferreira S., Fleischmann W.,
RA Flosser C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Houslin D., Houston K.A., Howland T.J., Wei M.H., Idegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kenison J.A., Ketchum K.A.,
RA Kimmel B.B., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laško P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Martel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mouton G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Murty D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
 RA Miranda M., Narusaka M., Nguyen M., Palm C.U., Sakurai T., Satou M.,
 RA Seki M., Shinn P., Southwick A., Shinzaki K., Davis R.W., Ecker J.R.,
 RA Theologis A.,
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC002387; AAB82636.2; -
 DR EMBL: AY024338; AAK00530.1; -
 DR EMBL: AY063864; AAL36220.1; -
 DR EMBL: AY117283; AAM51358.1; -
 DR PIR: E84888; E84888.
 DR GO: GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.
 DR GO: GO:0008233; F:peptidase activity; IEA.
 DR GO: GO:0008270; F:zinc ion binding; IEA.
 DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro: IPR009095; Peptidase_M22.
 DR InterPro: IPR009180; Pept_M22_Osialgl.
 DR Pfam: PF00814; Peptidase_M22; 1.
 DR PIRSF: PIRSF04537; Osialglc_ptds; 1.
 DR PRINTS: PR00789; OSIALOPTASE.
 DR ProDom: PD002367; Peptidase_M22; 1.
 DR TIGRFAMs: TIGR00329; gcp; 1.
 KM Complete proteome.
 SQ SEQUENCE 480 AA; 52995 MW; 20DD6A86AC1FPAD CRC64;

Query Match 31.5%; Score 668.5; DB 2; Length 480;
 Best Local Similarity 39.4%; Pred. No. 5.4e-43;
 Matches 154; Conservative 60; Mismatches 130; Indels 47; Gaps 7;

QY 38 IVLGIEISCDPTAAVNDETGNVIGEAIHSQTEVHLKKGIVPPAAQQLHRENIQRIYQ 97
 DB 85 VLDIEISCDPTAAVAVNGNBELISQVLSQAEILLVQGVAPKQAEHSHVIDKVVOD 144
 QY 98 ALSASGVSPDLISAATITIKPGLALSLGVGSFSLQVLGQLKPKPIPIHMEHAHALTR 157
 DB 145 ALDQANLTERKLSAVATITIGPGLSLCLRVGRKARVANGNSPLIVGCHHMAHALVRL 204
 QY 158 T-NKEPPEPVLVLLISGHCILLALVQGVSDFLILGKSLDAPGMDLKVARRSLIKHPKC 216
 DB 205 VEGELSPFPMALLISGHNLLVLAKHLGQYQLGTTVDDAIGEAVDKAKMLGLDMH--- 261
 QY 217 STMSGGKAIEHLAKQGRFHFDDIKRPLHAKKNCDSFPGLOHVTDKIMKKEKEGIEKG 276
 DB 262 --RSGGAVBEVLLEGDAKSKYKFNVMKXKDCNFSYAGLTKQVRLAEAKE----IDAK 315
 QY 277 QILSSA-----ADIAATVQHTWACHLVKTRTHALIFCKQRDLIPQNNALVASSGVA 328
 DB 316 CPVSSANTEDRRNRADIPASFORVAVLHBEKCEBAIDWALE--LEPSIKMVISGVA 372
 QY 329 SNFYIRBALILTNATQCTLLCPPEPLCTDNGIMIANNGIERLRAGLGLHDIGIRYEPK 387
 DB 373 SNKYVRLANNIVENKMLKLVCPPEPLCTDNGVMVAMWAGLEHFRVG-----RVDP 422
 QY 388 -----PKCPGLGVDISKEVGEA 403
 DB 423 PPRATPEDEYVYDRLPRWPLGEEYAKGRSEA 453

RESULT 13
 Q73H71 PRELIMINARY; PRT; 335 AA.
 AC Q73H71;
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE Peptidase, M22 family protein.
 GN OrderedLocustNames=MD0699;
 OS Rickettsia pipientis wmel.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Rickettsiaceae; Wolbachiae; Wolbachia.
 OX NCBI_TaxID=55077;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX PubMed=15024419;

RA Wu M., Sun L.V., Vamathavan J.J., Riegler M., Deboy R.T.,
 RA Brownlie J.C., McDraw E.A., Martin W., Esser C., Almadhinejad N.,
 RA Wiegand C., Madupu R., Beanan M.J., Brinkac L.M., Daugherty S.C.,
 RA Durkin A.S., Kolonay J.F., Nelson W.C., Mohamoud Y., Lee P.,
 RA Berry K.J., Young M.B., Ulteback T.R., Widman J.F., Nierman W.C.,
 RA Paulsen I.T., Nelson K.E., Tettein H., O'Neill S.L., Eisen J.A.,
 RT "Phylogenomics of the reproductive parasite Wolbachia pipientis wmel:
 RT a streamlined genome overrurn by mobile genetic elements."
 RL Plos Biol. 2:327-341(2004).
 DR EMBL: AE017258; AAS14395.1; -
 DR TIGR: WP0699; -
 DR GO: GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.
 DR GO: GO:0008270; F:zinc ion binding; IEA.
 DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro: IPR009095; Peptidase_M22.
 DR InterPro: IPR009180; Pept_M22_Osialgl.
 DR Pfam: PF00814; Peptidase_M22; 1.
 DR PIRSF: PIRSF04537; Osialglc_ptds; 1.
 DR PRINTS: PR00789; OSIALOPTASE.
 DR ProDom: PD002367; Peptidase_M22; 1.
 DR TIGRFAMs: TIGR00329; gcp; 1.
 KM Complete proteome.
 SQ SEQUENCE 335 AA; 36634 MW; F4914CDA58BC9390 CRC64;

Query Match 27.2%; Score 577.5; DB 2; Length 335;
 Best Local Similarity 36.3%; Pred. No. 3.6e-36;
 Matches 128; Conservative 68; Mismatches 130; Indels 27; Gaps 6;

QY 37 KIVLGIEISCDPTAAVNDETGNVIGEAIHSQTEVHLKKGIVPPAAQQLHRENIQRIYQ 96
 DB 2 KTLIAVETSCDPTAAVAVNSDKQVLAHEILSQAE-HKKGQVPIETASRAHMEHISGLIK 60
 QY 97 EALSASGVSPDLISAATITIKPGLALSLGVGSFSLQVLGQLKPKPIPIHMEHAHALTR 156
 DB 61 SAVESKNLNFCDMLAATISGRELIGLIVGTMAAIAHVAQPTIAVNHLEHALVLR 120
 QY 157 LTNKVEFPFVLVLLISGHCILLALVQGVSDFLILGKSLDAPGMDLKVARRSLIKHPKC 216
 DB 121 LHBEVKEPFLVLLISGHCQFLIADQVGYIKLGETLIDSLGSAFVKVAKMLGL----- 174
 QY 217 STMSGGKAIEHLAKQGRFHFDDIKRPLHAKKNCDSFPGLOHVTDKIMKKEKEGIEKG 276
 DB 175 -SYPGGPLELEKAKKNGTRFYLPRAMIKRSGCNFSFGIKTVAKMLVDELKMS----- 228
 QY 277 QILSSAADIAATVQHTWACHLVKTRTHALIFCKQRDLIPQNNALVASSGVA SNFYIRBA 336
 DB 229 ---QDVDCASFOECISIDILLDRVSNATIMESLMIKIND---FYITGVAAANNPLREK 282
 QY 337 LEILTNATQCTLLCPPEPLCTDNGIMIANNGIERLRAGLGLHDIGIRYEPK 389
 DB 283 LKQHN---LNIFFPPNDLCTDNALVGMGTGIERLOKNT-----IDPLNFAFR 327

RESULT 14
 Q92LH8 PRELIMINARY; PRT; 360 AA.
 AC Q92LH8;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE PROBABLE O-SIALOGLYCOPROTEIN ENDOPEPTIDASE (EC 3.4.24.57).
 GN ORFName=SKC03230;
 OS Rickettsia melloti (Sinorhizobium melloti).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
 OX NCBI_TaxID=382;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1021;
 RX MEDLINE=21396507; PubMed=11481430; DOI=10.1073/pnas.161294398;
 RA Capela D., Barloy-Hubler F., Gouzy J., Boche G., Ampe F., Bactut J.,
 RA Boistard P., Becker A., Bourtiry M., Cadieu B., Dreano S., Gloux S.,
 RA Godrie T., Goffeau A., Kahn D., Kles B., Lelaure V., Masny D.,

RA Pohl T., Portetelle D., Puehler A., Fumelle B., Ramsperger U.,
 RA Renard C., Thebault P., Vandenbol M., Weidner S., Gallibert F.,
 RT "Analysis of the chromosome sequence of the legume symbiont
 RT Sinorhizobium meliloti strain 1021."
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
 DR EMBL, AL591792; CAC47657.1; -.
 DR GO, GO:0016787; F:Hydrolase activity; IEA.
 DR GO, GO:0008450; F:O-6-ialoglycoprotein endopeptidase activity; IEA.
 DR GO, GO:0008270; F:zinc ion binding; IEA.
 DR GO, GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro, IPR000905; Peptidase_M22.
 DR InterPro, IPR009180; Pept_M22_Osialgl.
 DR Pfam, PF00814; Peptidase_M22_1.
 DR PIRSF, PIRSF004537; Osialglc_ptcds; 1.
 DR PRINTS, PR00789; OSIALOPTASB.
 DR Prodom, PD002367; Peptidase_M22; 1.
 DR TIGRFAMs, TIGR00329; gcp; 1.
 KW Complete proteome; Hydrolase.
 SQ SEQUENCE 360 AA; 37906 MW; A07F946AB562EA86 CRC64;

Query Match 26.8%; Score 568.5; DB 2; Length 360;
 Best Local Similarity 37.6%; Pred. No. 1.9e-35;
 Matches 143; Conservative 57; Mismatches 133; Indels 47; Gaps 9;

QY 39 VLGIETSCDDTAAV--DE--TGNVLGEAHSQTEVHLKTGIVPPAAQQLHRENIGRI 94
 DB 3 ILGIETSCDETAASVVALDEBGRRIIDGVVLSQLEHSAGVGVPEIARAHAHVEALDGI 62
 QY 95 VOELASAGVSPSDLSAIAITIKPLALSLVGLSFLQVQQLKKPFIPIHMEAHALT 154
 DB 63 IEEALLRAGVTLRIDAAVATSGPGLIGLVGMTAKIARATGKPLVAVNHLGHALT 122
 QY 155 IRLTNKVEPPLVLLISGGHCLLALVOGVDFLLGKSIDIAFGMLDKVARRISLIKHP 214
 DB 123 ARLTDLGSLFPLMLLVSGHQLILVKGVEYERNGTTIDDLGAEFDTAKMLGL-PPY 181
 QY 215 ECSTMSGKALIEHLAKQGNRFPHDIKPLHNAKNCDFSTGLQ-----HYTDKI 263
 DB 182 -----GGPVAERAAQAGNAERFDFPRPLVGDARLDFSGKLTAVRQAQSLGPTDGD 235
 QY 264 IMKKEKEBIEKQGLISSAADIATVQHTMACHLYKTRHAILFCRKORDLLPQNNNAVLYA 323
 DB 236 I-----ADVCAISFORAISRTLRDVRGKRRF-ADFAVDOPALV 276
 QY 324 SGVAASNYIRRAELITNAQCTLLCPPEPLCTDNGIMIANGIERLAGIILHIDIEG 383
 DB 277 AGVAANQTLRRITLQSLCDEHGFRFAPLQCTDNNAMIAWAGARLAGL----PADG 332
 QY 384 IRVEPKC--PLGVDISKEVG 401
 DB 333 LDAAPRSRWPLDSEKALIG 352

RESULT 15
 Q98E16 PRELIMINARY: PRT; 362 AA.
 AC Q98E16.
 DT 01-OCT-2001 (TREMBLrel. 18, Created)
 DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)
 DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
 DE O-6-ialoglycoprotein endopeptidase (Gcp).
 GN Ordered locus names: mlr4224;
 OS Rhizobium loti (Mesorhizobium loti).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Phyllobacteriaceae; Mesorhizobium.
 OC NCBI_TaxID=381;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MAFF303099;
 RX MEDLINE=21082930; PubMed=11214968;
 RA Kaneo T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
 RA Matenabe A., Idekawa K., Ishikawa K., Kawashima M., Kimura T.,
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,

RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tabata S.,
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 RT Mesorhizobium loti."
 RL DNA Res. 7:331-338(2000).
 DR EMBL, AP003003; BAB50932.1; -.
 DR GO, GO:0008450; F:O-6-ialoglycoprotein endopeptidase activity; IEA.
 DR GO, GO:0008270; F:zinc ion binding; IEA.
 DR GO, GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro, IPR000905; Peptidase_M22.
 DR InterPro, IPR009180; Pept_M22_Osialgl.
 DR Pfam, PF00814; Peptidase_M22_1.
 DR PIRSF, PIRSF004537; Osialglc_ptcds; 1.
 DR PRINTS, PR00789; OSIALOPTASB.
 DR Prodom, PD002367; Peptidase_M22; 1.
 DR TIGRFAMs, TIGR00329; gcp; 1.
 KW Complete proteome.
 SQ SEQUENCE 362 AA; 37345 MW; 8E4B11C45895D1D7 CRC64;

Query Match 26.5%; Score 564; DB 2; Length 362;
 Best Local Similarity 38.4%; Pred. No. 4.3e-35;
 Matches 145; Conservative 51; Mismatches 148; Indels 34; Gaps 10;

QY 39 VLGIETSCDDTAAV--DE--TGNVLGEAHSQTEVHLKTGIVPPAAQQLHRENIGRI 94
 DB 4 VLGIETSCDETAASVVALDEBGRRIIDGVVLSQLEHSAGVGVPEIARAHAHVEALDGI 63
 QY 95 VOELASAGVSPSDLSAIAITIKPLALSLVGLSFLQVQQLKKPFIPIHMEAHALT 154
 DB 64 VEAALDSTTLADIDALIAATAGPGLVGLVGLMTRAKAIAAAGPLVAVNHLGHALT 123
 QY 155 IRLTNKVEPPLVLLISGGHCLLALVOGVDFLLGKSIDIAFGMLDKVARRISLIKHP 214
 DB 124 ARLTDLGSLFPLMLLVSGHQLILVKGVEYERNGTTIDDLGAEFDTAKMLGL-PPY 182
 QY 215 ECSTMSGKALIEHLAKQGNRFPHDIKPLHNAKNCDFSTGLQ-----HYTDKI 263
 DB 183 -----GGPVAERAAQAGNAERFDFPRPLVGDARLDFSGKLTAVRQAQSLGPTDGD 235
 QY 275 K-QGLISSAADIATVQHTMACHLYKTRHAILFCRKORDLLP-QNNNAVLYASGVASNY 332
 DB 231 PLGDD--QDVADICASQAAVADALADRVSRLARFQ--TFPGTKPALVAVAGVAAHNT 286
 QY 333 IRRALIELITNAQCTLLCPPEPLCTDNGIMIANGIERLAGIILHIDIEGIRVEPKCPL 392
 DB 287 IKATLERLCAEAGFTVAPPLKCTDNNAMIAWAGARLAGL----QENGPFVPRSRW 343
 QY 393 GVDISKEVGEASIKVPOL 410
 DB 344 PLD-----SISAPMV 353

Search completed: November 10, 2005, 16:52:01
 Job time: 110.35 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2005, 15:51:52 ; Search time 33.3 Seconds
(without alignments)
928.069 Million cell updates/sec

Title: US-10-649-273-2
Perfect score: 2125
Sequence: 1 MLILTKTAGVFFPKSRKRVY.....DISKEVGASIKVPLQKMEI 414

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:
1: /cgn2_6/ptodata/1/iaa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS.COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/Backfilest1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2125	100.0	414	4	US-10-067-443-2 Sequence 2, Appli
2	2090.5	98.4	439	4	US-10-067-443-19 Sequence 19, Appl
3	1385	65.2	267	4	US-10-067-443-22 Sequence 22, Appl
4	634	29.8	463	4	US-10-067-443-3 Sequence 3, Appli
5	549	25.8	179	4	US-10-067-443-25 Sequence 25, Appl
6	524	24.7	421	4	US-10-067-443-4 Sequence 4, Appli
7	524	24.7	421	4	US-10-067-443-28 Sequence 28, Appl
8	501	23.6	350	4	US-09-540-236-2726 Sequence 2736, Ap
9	499.5	23.0	401	4	US-09-252-991A-17372 Sequence 17372, A
10	488	23.0	342	1	US-08-087-797-3 Sequence 3, Appli
11	482.5	22.7	357	4	US-09-543-681A-6513 Sequence 6513, Ap
12	475.5	22.4	325	1	US-08-087-797-2 Sequence 2, Appli
13	474.5	22.3	343	4	US-09-489-039A-9221 Sequence 9221, Ap
14	472	22.2	363	4	US-09-107-532A-6609 Sequence 6609, Ap
15	461	21.7	336	3	US-08-987-121A-4 Sequence 4, Appli
16	457	21.5	327	4	US-10-067-443-5 Sequence 5, Appli
17	457	21.5	335	3	US-08-961-083-52 Sequence 52, Appl
18	457	21.5	335	4	US-09-536-784-52 Sequence 52, Appl
19	455	21.4	336	3	US-09-066-512-2 Sequence 2, Appli
20	454	21.4	366	4	US-09-134-000C-4956 Sequence 4956, Ap
21	453	21.3	336	4	US-09-583-110-4857 Sequence 4857, Ap
22	448	21.1	336	4	US-09-107-433-4221 Sequence 4221, Ap
23	443	20.8	368	3	US-09-134-001C-3909 Sequence 3909, Ap
24	433	20.4	341	3	US-03-145-624-2 Sequence 2, Appli
25	406.5	19.1	344	4	US-09-198-452A-213 Sequence 213, App
26	406.5	19.1	360	4	US-09-438-185A-196 Sequence 196, App
27	401.5	18.9	344	4	US-09-602-777A-148 Sequence 148, App

28	362	17.0	292	4	US-09-724-623-81 Sequence 81, Appl
29	359	16.9	214	4	US-09-328-352-4609 Sequence 4609, Ap
30	338.5	15.9	340	4	US-10-067-443-6 Sequence 6, Appli
31	319	15.0	273	4	US-09-710-279-728 Sequence 728, App
32	204	9.6	187	4	US-09-248-796A-19489 Sequence 19489, A
33	181.5	8.5	143	4	US-09-328-352-4387 Sequence 4387, Ap
34	109	5.1	515	4	US-09-252-991A-33046 Sequence 33046, A
35	100	4.7	225	4	US-09-602-777A-144 Sequence 144, App
36	97.5	4.6	580	4	US-09-489-039A-12013 Sequence 12013, A
37	97	4.6	601	3	US-09-155-036-5 Sequence 5, Appli
38	97	4.6	601	4	US-09-866-307-5 Sequence 5, Appli
39	97	4.6	614	3	US-09-155-036-6 Sequence 6, Appli
40	97	4.6	614	4	US-09-866-307-6 Sequence 6, Appli
41	96	4.5	612	4	US-09-902-540-13545 Sequence 13545, A
42	94.5	4.4	480	4	US-09-583-110-5050 Sequence 5050, Ap
43	94.5	4.4	481	4	US-09-107-433-3197 Sequence 3197, Ap
44	93.5	4.4	294	4	US-09-107-532A-6933 Sequence 6933, Ap
45	93.5	4.4	579	4	US-09-543-681A-5019 Sequence 5019, Ap

ALIGNMENTS

```
RESULT 1
US-10-067-443-2
; Sequence 2, Application US/10067443
; Patent No. 6642041
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED I
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266, 518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282, 814
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-067-443-2
Query Match      100.0%; Score 2125; DB 4; Length 414;
Best Local Similarity 100.0%; Pred. No. 1.4e-234;
Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 MLILTKTAGVFFPKSRKRVYFLRSFNHPTGLFLKILVGLTETSCDDPRAAVDETGV 60
      |||||||
DB      1 MLILTKTAGVFFPKSRKRVYFLRSFNHPTGLFLKILVGLTETSCDDPRAAVDETGV 60
QY      61 LGEALHSQTEVHLKTCGIYPPAAOQLHRENIQRIQVEALASGVSPPSDSAITIKPGI 120
      |||||||
DB      61 LGEALHSQTEVHLKTCGIYPPAAOQLHRENIQRIQVEALASGVSPPSDSAITIKPGI 120
QY      121 ALSTGVGSLFSLQVLGKLPPIPIHMEAHALLTRLTNKKVEPPLVLLISGHCILALV 180
      |||||||
DB      121 ALSTGVGSLFSLQVLGKLPPIPIHMEAHALLTRLTNKKVEPPLVLLISGHCILALV 180
QY      122 ALSTGVGSLFSLQVLGKLPPIPIHMEAHALLTRLTNKKVEPPLVLLISGHCILALV 180
      |||||||
DB      122 ALSTGVGSLFSLQVLGKLPPIPIHMEAHALLTRLTNKKVEPPLVLLISGHCILALV 180
QY      181 QGVSPFLILGKSLDIAPGMDLKVARRSLIHPBCSTWSGGKAIHLAKQGRFFIDIK 240
      |||||||
DB      181 QGVSPFLILGKSLDIAPGMDLKVARRSLIHPBCSTWSGGKAIHLAKQGRFFIDIK 240
QY      241 PHLHAKNDPFTGLQHTYTDKTIKKKEKRGILGQILSSAADIAATVOHTMACILVXR 300
      |||||||
DB      241 PHLHAKNDPFTGLQHTYTDKTIKKKEKRGILGQILSSAADIAATVOHTMACILVXR 300
QY      301 THRALLPCKQRRLLPONNAVLVAAGVANSFYIRRLALITLTVNAOCTLLCPPRLCTDNG 360
      |||||||
DB      301 THRALLPCKQRRLLPONNAVLVAAGVANSFYIRRLALITLTVNAOCTLLCPPRLCTDNG 360
```

QY 361 IMIANGIERLRAGIGIILHDIIEGIRYBPKPLGVDSKEVGEASIKVPOLKMEI 414
 Db 361 IMIANGIERLRAGIGIILHDIIEGIRYBPKPLGVDSKEVGEASIKVPOLKMEI 414

RESULT 2

US-10-067-443-19
 ; Sequence 19, Application US/10067443

Patent No. 6642041

GENERAL INFORMATION:

APPLICANT: Bristol-Myers Squibb Company

TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED

FILE REFERENCE: D0073 NP

CURRENT APPLICATION NUMBER: US/10/067,443

CURRENT FILING DATE: 2002-02-05

PRIOR APPLICATION NUMBER: US 60/266,518

PRIOR FILING DATE: 2001-02-05

PRIOR APPLICATION NUMBER: US 60/282,814

PRIOR FILING DATE: 2001-04-10

NUMBER OF SEQ ID NOS: 71

SOFTWARE: Patentin version 3.0

SEQ ID NO 19

LENGTH: 439

TYPE: PRT

ORGANISM: homo sapiens

US-10-067-443-19

Query Match 98.4%; Score 2090.5; DB 4; Length 439;

Best Local Similarity 93.8%; Pred. No. 1.4e-230;

Matches 412; Conservative 0; Mismatches 2; Indels 25; Gaps 1;

QY 1 MLITKTVGVFPKPKSKRYEFLRSFNFPGTLFLHKIYLGIFETSCDDTAAAVNDGTGV 60
 Db 1 MLITKTVGVFPKPKSKRYEFLRSFNFPGTLFLHKIYLGIFETSCDDTAAAVNDGTGV 60
 QY 61 LGEATHSQTEVHLKTGGIVPPAAQOLHRENIQRIVOEALISASGVSPSDISATITIKPGL 120
 Db 61 LGEATHSQTEVHLKTGGIVPPAAQOLHRENIQRIVOEALISASGVSPSDISATITIKPGL 120
 QY 121 ALSIGVGSFSLQVLGOLKKPFIPIHMEBAHALTIRLTNKVEFPFLVLLISGHCILALV 180
 Db 121 ALSIGVGSFSLQVLGOLKKPFIPIHMEBAHALTIRLTNKVEFPFLVLLISGHCILALV 180
 QY 181 QGVSDPFLLGKSLDIAPGMDLVARRLSLIYHPECSTWSGGKAIHIAKQNRFFHDIX 240
 Db 181 QGVSDPFLLGKSLDIAPGMDLVARRLSLIYHPECSTWSGGKAIHIAKQNRFFHDIX 240
 QY 241 PRLHAKKCDPSFTGLQHTVDKTIIMKKEKEBGI-----RK 275
 Db 241 PRLHAKKCDPSFTGLQHTVDKTIIMKKEKEBGI-----RK 275
 QY 276 GQISSADIAATVOHTMAACHVKRTHRAILFCQKQDILLPONNAVLVASGVASNFYIR 335
 Db 301 GQISSADIAATVOHTMAACHVKRTHRAILFCQKQDILLPONNAVLVASGVASNFYIR 360
 QY 336 ALBITTNAQTCLLCPPLRCTDNGIMIAMNGIERLRAGIGIILHDIIEGIRYBPKPLGV 395
 Db 361 ALBITTNAQTCLLCPPLRCTDNGIMIAMNGIERLRAGIGIILHDIIEGIRYBPKPLGV 420
 QY 396 ISKEVGEASIKVPOLKMEI 414
 Db 421 ISKEVGEASIKVPOLKMEI 439

RESULT 3

US-10-067-443-22

; Sequence 22, Application US/10067443

Patent No. 6642041

GENERAL INFORMATION:

APPLICANT: Bristol-Myers Squibb Company

TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED

FILE REFERENCE: D0073 NP
 ; CURRENT APPLICATION NUMBER: US/10/067,443
 ; CURRENT FILING DATE: 2002-02-05
 ; PRIOR APPLICATION NUMBER: US 60/266,518
 ; PRIOR FILING DATE: 2001-02-05
 ; PRIOR APPLICATION NUMBER: US 60/282,814
 ; PRIOR FILING DATE: 2001-04-10
 ; NUMBER OF SEQ ID NOS: 71
 ; SOFTWARE: Patentin version 3.0
 ; SEQ ID NO 22
 ; LENGTH: 267
 ; TYPE: PRT
 ; ORGANISM: homo sapiens
 ; US-10-067-443-22

Query Match 65.2%; Score 1385; DB 4; Length 267;

Best Local Similarity 100.0%; Pred. No. 5.1e-150;

Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 148 MEHAALTIRLTNKVEFPFLVLLISGHCILALVQGVSDPFLLGKSLDIAPGMDLVARR 207
 Db 1 MEHAALTIRLTNKVEFPFLVLLISGHCILALVQGVSDPFLLGKSLDIAPGMDLVARR 60
 QY 208 LSLIKHPECSTWSGGKAIHIAKQNRFFHDIXPPLHAKKCDPSFTGLQHTVDKTIIMK 267
 Db 61 LSLIKHPECSTWSGGKAIHIAKQNRFFHDIXPPLHAKKCDPSFTGLQHTVDKTIIMK 120
 QY 268 EKEBGEIKQIISADIAATVOHTMAACHVKRTHRAILFCQKQDILLPONNAVLVASGV 327
 Db 121 EKEBGEIKQIISADIAATVOHTMAACHVKRTHRAILFCQKQDILLPONNAVLVASGV 180
 QY 328 ASNFYIRRALBITTNAQTCLLCPPLRCTDNGIMIAMNGIERLRAGIGIILHDIIEGIR 387
 Db 181 ASNFYIRRALBITTNAQTCLLCPPLRCTDNGIMIAMNGIERLRAGIGIILHDIIEGIR 240
 QY 388 PKCPLGVDSKEVGEASIKVPOLKMEI 414
 Db 241 PKCPLGVDSKEVGEASIKVPOLKMEI 267

RESULT 4

US-10-067-443-3

; Sequence 3, Application US/10067443

Patent No. 6642041

GENERAL INFORMATION:

APPLICANT: Bristol-Myers Squibb Company

TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED

FILE REFERENCE: D0073 NP

CURRENT APPLICATION NUMBER: US/10/067,443

CURRENT FILING DATE: 2002-02-05

PRIOR APPLICATION NUMBER: US 60/266,518

PRIOR FILING DATE: 2001-02-05

PRIOR APPLICATION NUMBER: US 60/282,814

PRIOR FILING DATE: 2001-04-10

NUMBER OF SEQ ID NOS: 71

SOFTWARE: Patentin version 3.0

SEQ ID NO 3

LENGTH: 463

TYPE: PRT

ORGANISM: Arabidopsis thaliana

US-10-067-443-3

Query Match 29.8%; Score 634; DB 4; Length 463;

Best Local Similarity 37.9%; Pred. No. 1.8e-63;

Matches 145; Conservative 61; Mismatches 129; Indels 48; Gaps 7;

QY 38 IVIGITSCDDTAAAVYDGTGVNVLGEAIIHQTVEHLKTGGIVPPAAQOLHRENIQRIVOE 97
 Db 85 IVIGITSCDDTAAAVYDGTGVNVLGEAIIHQTVEHLKTGGIVPPAAQOLHRENIQRIVOE 141
 QY 98 ALSASGVSPSDISAIATITIKRGLALSLGVGLSFLQVLGQKKPFIPIHMEBAHALTIRL 157

```
Db 142 ALDKNLPEKDISAVALTIGPGLSLCLRGVKKARKVAANFSLPIYGVHMEHALVARI 201
Qy 158 T-NKVEPFLVLLISGHCCLLALVOGVSDFLLGSLDIAPGMDLKVARRSLIHKPRC 216
Db 202 VEGELSPFFMALLISGHNLLVLAHKLGYTOLGTTVDALIGAPKTKAWGLDMH--- 258
Qy 217 STMSGKAIIEHLAKQGNRFHEDIKPLHHAKNCDPSFTGLQHYTDKIIMKKEKEGIEKG 276
Db 259 --RSGGPAYEELEALGDAKSVKFNVPKHYKDCNFSYALKTQVRLAIBAKE----- 308
Qy 277 QILSSAALIAATVQHTMACHLVKTRHRAIFCKQBDLPONNAVIVASGVASNFYIRRA 336
Db 309 --INRRADIAASFQVAVVLALEEKCEKRAIDWALE--LEPSIKHWISGVASNKYVRLR 363
Qy 337 LEITNATQCTLCPPLRLCTDNGIMIAMNGIERLAPGLIHDEIGIYE----- 387
Db 364 LNNIVENKULVCPSPSLCTDNGVMAVTGIEHFRVG-----RYDEPPRATPE 413
Qy 388 -----PKCPGLVDISKEVGEA 403
Db 414 DYVYDLRRWPLGEEYAKGRSEA 436
```

RESULT 5

```
US-10-067-443-25
; Sequence 25, Application US/10067443
```

```
; Patent No. 6642041
```

```
; GENERAL INFORMATION:
```

```
; APPLICANT: Bristol-Myers Squibb Company
```

```
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE HIGHLY EXPRESSED
```

```
; FILE REFERENCE: D0073 NP
```

```
; CURRENT APPLICATION NUMBER: US/10/067,443
```

```
; PRIOR FILING DATE: 2002-02-05
```

```
; PRIOR APPLICATION NUMBER: US 60/266,518
```

```
; PRIOR FILING DATE: 2001-02-05
```

```
; PRIOR APPLICATION NUMBER: US 60/282,814
```

```
; NUMBER OF SEQ ID NOS: 71
```

```
; SOFTWARE: PatentIn version 3.0
```

```
; SEQ ID NO 25
```

```
; LENGTH: 179
```

```
; TYPE: PRT
```

```
; ORGANISM: homo sapiens
```

```
US-10-067-443-25
```

```
Query Match 25.8%; Score 549; DB 4; Length 179;
Best Local Similarity 66.5%; Pred. No. 2e-54;
```

```
Matches 119; Conservative 0; Mismatches 0; Indels 60; Gaps 1;
```

```
Qy 38 IVLGIEFSCDDTAAAVDETGNVGEALHSQTEVHLKTGGIYPPAAQQLHRENIQRIYOE 97
Db 1 IVLGIEFSCDDTAAAVDETGNVGEALHSQTEVHLKTGGIYPPAAQQLHRENIQRIYOE 60
Qy 98 ALSASGVSPDLSAIIATTIKPGLALSLGVGSFSLQVGLKKPPIPCCATTCATCATAT 144
Db 61 ALSASGVSPDLSAIIATTIKPGLALSLGVGSFSLQVGLKKPPIPCCATTCATCATAT 120
Qy 145 -----IHHMEAHALTIR 156
Db 121 GGAGGCTCAGCAGCTACTATTAGCTTGACCAATAAGTAGAATTCIHHEAHALTIR 179

RESULT 6
US-10-067-443-4
; Sequence 4, Application US/10067443
; Patent No. 6642041
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE HIGHLY EXPRESSED
; FILE REFERENCE: D0073 NP
```

```
; CURRENT APPLICATION NUMBER: US/10/067,443
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 421
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-067-443-4
```

```
Query Match 24.7%; Score 524; DB 4; Length 421;
```

```
Best Local Similarity 33.4%; Pred. No. 6.3e-51;
```

```
Matches 130; Conservative 77; Mismatches 146; Indels 36; Gaps 8;
```

```
Qy 39 VLGIEFSCDDTAAAVDETGNVGEALHSQTEVHLKTGGIYPPAAQQLHRENIQRIYOE 98
Db 25 VLGIEFSCDDTAAVAIVNEKREILSSEYTERAIQROGGGINPSCALQHRENIPLRIEKC 84
Qy 99 LSASGVSPDLSAIIATTIKPGLALSLGVGSFSLQVGLKKPPIPIHHMEAHALTIRLT 158
Db 85 LNDAGTSPKDLDAVAVTVTGLVIALKEGISAIGFAKGRPLPIVHHRARALSTLLV 144
Qy 159 -NKVEPFLVLLISGHCCLLALVOGVSDFLLGSLDIAPGMDLKVARRSLIHKPRCS 217
Db 145 DSVRRPFSVAVLISGHALISVADYBKFLYQOSVSGSECTDKVARKQGLD-GSEFD 203
Qy 218 TMSGKAIIEHLAKQGN--RFHEDIKPLHHAKNCDPSFTGLQHYTDKIIMKKEKEGIE 274
Db 204 GHVGAAVEILASRAADGHLRYPIFLPNVPKANNMFDQIKSYLMLERLRKNSERSID 263
Qy 275 KQQLSSAADIAATVQHTMACHLVKTRHRAIFCKQBDLPONNAVIVASGVASNFYIR 334
Db 264 -----IDPFCASLONTVAHHSKLIHFFESLSIEKPKO---LVIGGVAAHQYIF 313
Qy 335 RALEITNATQCTLCPPLRLCTDNGIMIAMNGIERLAPGLIHDEIGIYRYP----- 388
Db 314 GAISKLSAANVTYITVLLSLCTDNAMETAYSQL-----LMLVNSEAIWMPNDIPDT 367
Qy 389 -----KCPGLVDISKEVGEASIKVPOLKM 412
Db 368 IYAHARSDIGTDASSRI-----IDTPRRKL 392
```

RESULT 7

```
US-10-067-443-28
; Sequence 28, Application US/10067443
```

```
; Patent No. 6642041
```

```
; GENERAL INFORMATION:
```

```
; APPLICANT: Bristol-Myers Squibb Company
```

```
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE HIGHLY EXPRESSED
```

```
; FILE REFERENCE: D0073 NP
```

```
; CURRENT APPLICATION NUMBER: US/10/067,443
```

```
; PRIOR FILING DATE: 2002-02-05
```

```
; PRIOR APPLICATION NUMBER: US 60/266,518
```

```
; PRIOR FILING DATE: 2001-02-05
```

```
; PRIOR APPLICATION NUMBER: US 60/282,814
```

```
; NUMBER OF SEQ ID NOS: 71
```

```
; SOFTWARE: PatentIn version 3.0
```

```
; SEQ ID NO 28
```

```
; LENGTH: 421
```

```
; TYPE: PRT
```

```
; ORGANISM: Caenorhabditis elegans
```

```
US-10-067-443-28
```

```
Query Match 24.7%; Score 524; DB 4; Length 421;
```

```
Best Local Similarity 33.4%; Pred. No. 6.3e-51;
```

```
Matches 130; Conservative 77; Mismatches 146; Indels 36; Gaps 8;
```



```

QY 39 VLGIETSCDDTAAAVVDETVGLGEAHSQTEVHLKTCGIIVPAAQOLHRENIQRIYVEA 98
DB 25 VLGIETSCDDTAAVAVNKEKREILSERTERALICQOQGINSVACLOHRENLPLIEKIC 84
QY 99 LLSAGVSPSDLSAIAATTIKPGIALSLGVLSFSLQVLQKKPPIPIHMEHAALTIRLT 158
DB 85 LNDAGTSPKDLDAVAVTVPGLVIALKEGISALIFAKKHRLPLIPVHMRHALSTILLY 144
QY 159 -NKVEPFLVLLISGGHCLLALVQGVSPDLLGKSLDIPAGMDLKVARRLSLIGHPES 217
DB 145 DDSVRFPPSAVLLSGGHALISVABDEVKFLYGVGSVSPGECIDKVAQLDL- GSEPD 203
QY 218 TMSGKAIEHLAKQGN---RFPDIPKPLIHAKNCDPFTGQHTYDKIMKEKEEGIE 274
DB 204 GIHVGAIVEILASRSADGHLRYFELNVPKANNFQOIKSYANLLEKLNKSETSID 263
QY 275 KGOILSSADIAATVOHTMACHLVKTRHAILFCQORDLLPQNNAVLVASGVASNFYIR 334
DB 264 -----IPDCASLQNTVAHHSKLIHFFESLSEQELPRQ---LVIGGVAAHQYIF 313
QY 335 RALELTNATQCTLLCPPRLCTDNGIMIANNGIRLAKGLIHDIGIRYEP----- 388
DB 314 GAISRLSAHVVTIKVLLSLCTDVAEMIAVSGL-----LMLVNRSEAIWVRPNDIPDT 367
QY 389 -----KCPIDVDSKEVGEASIKVPLKX 412
DB 368 IYAHASDIDGTASSEI-----IDTPRKX 392

```

RESULT 8

```

US-09-540-236-2726
; Sequence 2726, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATAR
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2005-001
; CURRENT APPLICATION NUMBER: US/09/540,236
; NUMBER OF SEQ ID NOS: 2000-04-04
; SEQ ID NO 2726
; LENGTH: 350
; TYPE: PRT
; ORGANISM: M.catarrhalis
US-09-540-236-2726

```

Query Match 23.6%; Score 501; DB 4; Length 350;

Best Local Similarity 34.5%; Pred. No. 2e-48; Indels 44; Gaps 8;

Matches 126; Conservative 56; Mismatches 139;

```

QY 39 VLGIETSCDDTAAAVVDETVGLGEAHSQTEVHLKTCGIIVPAAQOLHRENIQRI 94
DB 6 VLGIETSCDDTAAVAVNKEKREILSERTERALICQOQGINSVACLOHRENLPLIE 65
QY 95 VQELASAGVSPSDLSAIAATTIKPGIALSLGVLSFSLQVLQKKPPIPIHMEHAALT 154
DB 66 FNEMDQANITKSIDAVATYKPGILGALMTGALFGRTLAAYGLGAPVGVHMEGHLLA 125
QY 155 IRLT--NKVEPFLVLLISGGHCLLALVQGVSPDLLGKSLDIPAGMDLKVARRLSL 212
DB 126 PLLASDDPPSPFVCLLVSGGHTMLVRADGVAYQLLSIDAVGECFDKTKAKMLKL-P 184
QY 213 HPECSTMGSKAIEHLAKQGNRFPDIPKPLIHAKNCDPFTGLOHVTYDKIMKEKEEG 272
DB 185 YP-----GGPIETKLAKNGNPHAYELRPPMOH-KGLDPSFGSKMAIHNILKDTNMA 237
QY 273 IEKQILSSADIAATVOHTMACHLVKTRHAILFCQORDLLPQNNAVLVASGVASNFY 332
DB 228 DP-----ATRADIASFEYAVVDTLVKKCTKALQMTGIRQ-----LVVAGGVASNGT 284
QY 333 IRRALIELTNATQCTLLCPPRLCTDNGIMIANNGIERLRAG-----L 375

```

```

DB 285 IRRTLTETLRQIDASVYAPTELCTDNGIMIAVAGCFRLSKQSDDLAVRCIPRMDTML 344
QY 376 GILHD 380
DB 345 GLEVD 349

```

RESULT 9

```

US-09-252-991A-17372
; Sequence 17372, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfeld et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: ABRUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 17372
; LENGTH: 401
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17372

```

Query Match 23.5%; Score 499.5; DB 4; Length 401;

Best Local Similarity 34.4%; Pred. No. 3.8e-48;

Matches 133; Conservative 67; Mismatches 158; Indels 29; Gaps 10;

```

QY 13 KPSKRVYEPFLKSNFHPGTLFLHKI-----VLGIETSCDDTAAAVVDETVGLGEAHS 66
DB 31 KASDRKTLVWRKADYHARLITCOALRKPMRVTLGIETSCDDTAAVAVNKEKREILSER 90
QY 67 SQTEVHLKTCGIIVPAAQOLHRENIQRIYVEA LLSAGVSPSDLSAIAATTIKPGIALSLGV 126
DB 91 SQDLHRVYGVGVPELASHDVRMLPLIRQVLDSGCTPADIDALAYTAPGLVGLALV 150
QY 127 GLSFSLQVLQKKPPIPIHMEHAALTIRLTNKK-VEPFLVLLISGGHCLLALVQGVSD 185
DB 151 GASCAQAMAFAMKVPVAVGVHMEGHLLAPLLEQPPRFVVALVSGHTQLVAVDSIGR 210
QY 186 FULLGKSLDIPAGMDLKVARRLSLKHPECSTMGSKAIEHLAKQGNRFPDIPKPLH 245
DB 211 YQLGSSVDDDAAGEAFDXTAKLIGL-GYP-----GGPEIARLBERGTPGRFVFPFRMTD 263
QY 246 AKNCDFSSFTGLQHVYTDKIMKEKEEGIEKQILSSADIAATVOHTMACHLVKTRHAI 305
DB 264 RPLGIDPSFSLKTFPTN-TWQRCVEAGDSEQ---TRCDIALAFQAVAVETLLIKCRAL 319
QY 306 LFCQORDLLPQNNAVLVASGVASNFYIRALIELTNATQCTLLCPPRLCTDNGIMIAN 365
DB 320 ---KQGL-KN--LVVAGGVASANOALSGLEKMLGEMKGOVFYARPRCTDNGIMIA 371
QY 366 NGIERLRAGVILHDIGIRYEPKPL 392
DB 372 AGCGRLLAG---QHDGPAISVQPRWPM 395

```

RESULT 10

```

US-08-087-797-3
; Sequence 3, Application US/08087797
; Patent No. 5543312
; GENERAL INFORMATION:
; APPLICANT: Mellore, Alan
; APPLICANT: Lo, Reggie Y. C.
; APPLICANT: Abdullah, Khalid M.
; TITLE OF INVENTION: Pasteurella Haemolytica
; TITLE OF INVENTION: Glycoprotease

```


QY 97 EALSASGVSFSDLSAATTIKRGALSLGVLSFSLQVLGQKPFPIPHMEALTR 156
DB 89 DALVEAGVSABDSVAATVGPGLVSLIGISAAPAMAHQPLIPVNHAGHIYAR 148
QY 157 LTNKVEFPFLVLLISGCHCLALVGVSDPFLIGKSLDIPAGMDLKVARRLSLIKHPCC 216
DB 149 LVKFPQFPLMLLVSGGHTLVYMEDESGYEIIGSTRDDAGEAIDKGRVGL----- 202
QY 217 STMSGKAIIEHLAKQG-NRFPDIKPLPHAKNCDFSFTGLQHTVDKIIMK-KEKEBEGIE 274
DB 203 -SYPSGKEIDDLAQHGKDNHYH--PRAMIHEDNDVDFSGLSKSAFINLVHNAQORGEDDD 259
QY 275 KGQLLSADIAATVQHTMACHLVKTRAILFCQKQBDLIPONNAV--LVASGVASNFY 332
DB 260 KN-----DLAASFQASVIDVLITKTLRA---C-----QNVYKQLVVAGVAAQ 302
QY 333 IRRALIELTNA--TQCTLLCPPRLCTDNGIMI 363
DB 303 LREGLQALSAKLPEVELVTPPLRLCGDMAAMI 335

RESULT 15

US-08-987-121A-4
; Sequence 4, Application US/08987121A
; GENERAL INFORMATION:
; APPLICANT: Hoskins, Jo Ann
; APPLICANT: Tang, Joseph Chiou-Chung
; APPLICANT: Treadway, Patti Jean
; TITLE OF INVENTION: Streptococcus Pneumoniae Gene Sequence
; TITLE OF INVENTION: Gcp
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: U.S.
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/987,121A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Webster, Thomas D.
; REGISTRATION NUMBER: 39,872
; REFERENCE/DOCKET NUMBER: X-
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-3334
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 336 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-987-121A-4

Query Match 21.7%; Score 461; DB 3; Length 336;

Best Local Similarity 34.5%; Pred. No. 7.3e-44;

Matches 116; Conservative 58; Mismatches 132; Indels 30; Gaps 8;

QY 37 KIVIGIETSCDDTAADVDEGTGVGBAISHQTEVHLKTGGLVPPAAQOLHRENIQRIYQ 96
DB 4 KYILAFETSCDBTSVAIVAKNDELLSNVIAQIESHKRFGGVPEVVASRHHVEVITACIE 63
QY 97 EALSASGVSFSDLSAATTIKRGALSLGVLSFSLQVLGQKPFPIPHMEALTR 156
DB 64 EALAEAGITBEDVAVAVTYGPGLVGALLVGSAAKAPAMAHGFLIPVNHAGHLMMAQ 123

QY 157 LTNKVEFPFLVLLISGCHCLALVGVSDPFLIGKSLDIPAGMDLKVARRLSLIKHPCC 216
DB 124 SVEPLEFPFLALIVSGGHTLVVVSAGDYKIVGETRDDAVGEAYDKGRVMGI----- 177
QY 217 STMSGKAIIEHLAKQGNRFHFDIKPLPHAKNCDFSFTGLQHTVDKIIMKKEBEGIEKG 276
DB 178 -TYPAGREIDELAHQHD1-YDPPRAMIKEDNDLSEFSGLKSAFINLHNAE-----QKG 230
QY 277 QILSSADIAATVQHTMACHLVKTRAILFCQKQBDLIPONNAVLVASGVASNFYIRRA 336
DB 231 ESIST-EDLCASFOAAVMDILMAKTKAL-----EKYPK--TLVAGGVAAANKGILRR 281
QY 337 LEILTNAQTCTLLCPPRLCTDNGIMI-----KN 366
DB 282 L--ATEITDVNVIIIPPLRLCGDMAAMIAYASVSENN 315

Search completed: November 10, 2005, 16:54:34
Job time : 35.3 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 10, 2005, 16:59:00 ; Search time 1374.75 Seconds
(without alignments)
2490.420 Million cell updates/sec

Title: US-10-649-273-2
Perfect score: 2125
Sequence: 1 MLILTKTAGVFPPKSKRRVY.....DISKEVGASIVPOLKMEI 414

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 9794790 seqs, 413490567 residues

Total number of hits satisfying chosen parameters: 19589580

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2_1/USPRO.spool/p/US10649273/rnatc_02112005_091339_15609/app_query.fasta_1.1429
-DB=published Applications_NA -QPM=faetap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=numan40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=-1 -MODE=LOCAL -OUTFMT=pro -NORM=ext -HEARSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10649273 @CCN 1 1 1034 @rnatc_02112005_091339_15609
-NCPU=6 -ICPU=3 -NO MMAP -LARGEBUQUER -NEG SCORES=0 -WAIT -DSPELOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications_NA:*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:*
19: /cgn2_6/ptodata/1/pubpna/US10G_PUBCOMB.seq:*
20: /cgn2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq:*
21: /cgn2_6/ptodata/1/pubpna/US10I_PUBCOMB.seq:*
22: /cgn2_6/ptodata/1/pubpna/US10J_PUBCOMB.seq:*
23: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
24: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
25: /cgn2_6/ptodata/1/pubpna/US11A_PUBCOMB.seq:*
26: /cgn2_6/ptodata/1/pubpna/US11B_NEW_PUB.seq:*
27: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
28: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2125	100.0	2197	15	US-10-067-443-1
2	2125	100.0	2197	19	US-10-649-273-1
3	2125	100.0	2197	19	US-10-651-722-1
4	2125	100.0	2572	22	US-10-480-988-36
5	2090.5	98.4	1387	15	US-10-067-443-21
6	2090.5	98.4	1387	19	US-10-649-273-21
7	2090.5	98.4	1387	19	US-10-651-722-21
8	2090.5	98.4	1387	22	US-10-887-553A-1047
9	2088	98.3	1245	15	US-10-012-140-6
10	2088	98.3	1820	15	US-10-012-140-4
11	1944	91.5	2208	18	US-10-094-749-400
12	1944	91.5	2890	21	US-10-723-860-7447
13	1747	82.2	1416	18	US-10-120-988-177
14	1385	65.2	1526	15	US-10-067-443-23
15	1385	65.2	1526	19	US-10-649-273-23
16	1385	65.2	1526	19	US-10-651-722-23
17	1186.5	55.8	14364	15	US-10-067-443-20
18	1186.5	55.8	14364	19	US-10-649-273-20
19	1186.5	55.8	14364	19	US-10-651-722-20
20	714.5	33.6	1601	26	US-11-097-143-34190
21	687.5	32.4	3656	26	US-11-097-143-34189
22	682	32.1	1917	19	US-10-424-599-66417
23	599	28.2	2734	24	US-10-450-763-20426
24	578	27.2	1628	20	US-10-437-963-11249
25	531.5	25.0	4360	15	US-10-081-051-2
26	524	24.7	1146	18	US-10-282-122A-14674
27	502	23.6	94750	19	US-10-672-787-38
28	502	23.6	1032	18	US-10-282-122A-31043
29	494.5	23.3	1026	9	US-09-815-242-7701
30	492.5	23.2	1026	18	US-10-282-122A-30016
31	492.5	23.2	1026	18	US-10-282-122A-30016
32	492.5	23.2	1026	24	US-10-958-216-49
33	490.5	23.0	1026	24	US-10-958-216-51
34	488.5	23.0	1194	21	US-10-425-115-167432
35	484.5	22.8	1014	9	US-09-815-242-6207
36	484.5	22.8	1014	18	US-10-282-122A-20511
37	484.5	22.8	1014	18	US-10-282-122A-39301
38	483.5	22.8	1014	18	US-10-282-122A-41977
39	482.5	22.7	1020	18	US-10-282-122A-31254
40	481.5	22.4	1014	9	US-09-815-242-9682
41	475.5	22.4	1029	9	US-09-815-242-6946
42	475.5	22.4	1029	18	US-10-282-122A-22020
43	475.5	22.4	1830121	18	US-10-329-670-1
44	475.5	22.4	1830121	21	US-10-158-865-1
45	475.5	22.4	1830121	24	US-10-981-687-1

ALIGNMENTS

RESULT 1
US-10-067-443-1
; Sequence 1, Application US/10067443
; Publication NO. US20030082782A1
; APPROPRIATE INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN
; TITLE OF INVENTION: SPINAL CORD, MP-1
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067, 443
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266, 518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282, 814

PRIOR FILING DATE: 2001-04-10
NUMBER OF SEQ ID NOS: 71
SOFTWARE: PatentIn version 3.0
SEQ ID NO 1
LENGTH: 2197
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (231)..(1472)
US-10-067-443-1

Alignment Scores:

Pred. No.:	3,35e-247	Length:	2197
Score:	2125.00	Matches:	414
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	15	Gaps:	0

US-10-649-273-2 (1-414) x US-10-067-443-1 (1-2197)

```
QY      1 MetLeuIleLeuThrIlySthraIagIValIphePheLySProSeLySArgLySValIYr 20
        |||
DB      231 ATGCTAATCTTGACTAGACTGACAGAGATTTTAAACATCAAAAGAAAGTTAT 230
QY      21 GIUpheLeuArgSerPheAsnPhenIspProGIYThLeuPheLeuHIslySIIleValIeu 40
        |||
DB      231 GAATTTTAAAGAGATTAAATTTTCCTCTGAAACCTATTCTTCAATAAATAGATTG 350
QY      41 GIYIleGIuThSerCySaspSPThraIaAlaIValIAspGIuThrGIYAsnVal 60
        |||
DB      351 GGAAATTGAACCTAGTTGTATGATACACAGAGCTGCTGTGTGATGAAACTGAAATGTG 410
QY      61 LeuGIYGIuValaIIehIsSerGIuThrGIuValaIIeLeuLySthrGIYIIleValIPro 80
        |||
DB      411 TTGGAGAAAGCAATACATTCCTCCAAACTGAAGTTCATTAAACAGGTGGATTTGCTT 470
QY      81 ProAlaIaGIuGIuLeuHIsArgGIuAenIIeGIuArgIIleValGIuGIuAlaLeuSer 100
        |||
DB      471 CCAGACGCTCAACAGCTTCACAGAGAAATATTCACAGATAGTACAAAGAGCTCTTCT 530
QY      101 AlaSerGIYValSerProSerAspLeuSerAlaIIaIaIaThrThrIIeLySProGIYLeu 120
        |||
DB      531 GCCAGTGAAGTCTCTCCAGAGTACCTCTCCAGCAATTCGCACTCCATTAACCAAGACTT 590
QY      121 AlaLeuSerLeuGIYValGIYLeuSerPheSerLeuGIuLeuValGIYGIuLeuLyS 140
        |||
DB      591 GCTTTTAGCTGGAGAGTGGCTTATCATTTAGCTTACAGCTGGTAGGACAGTTAAAAAG 650
QY      141 PropheIIeProIIeHIsHISmetGIuAlaHIsAlaLeuThrIIeArgLeuThrAsnLyS 160
        |||
DB      651 CCATTCATTCCTCATTCATATGAGAGGCTCATGCACTTCTATTAGTGTGACCAATAAA 710
QY      161 ValGIuPheProPheLeuValIleuLeuIIeSerGIYHIsCySleuLeuAlaLeuVal 180
        |||
DB      711 GTAGAAATTTCTTTTATGTTCTTTGATTTCTGGAGGTACTGTCTGTGGCACTTAGTT 770
QY      181 GIuGIYValSerAspPheLeuLeuLeuGIYLySerLeuAspIIleAlaProGIYAspMet 200
        |||
DB      771 CAAGAGATTTCACATTTTCTGCTTCTTGAAAAGTCTTTTGACATAGACAGAGTGACATG 830
QY      201 LeuAspLySValaIIaArgArgLeuSerLeuIIeLySIIspProGIYCySserThMetSer 220
        |||
DB      831 CTTGACAAAGTGGCAAGAGACTTTCTTTATATAAACATCCAGAGTCTCCACCATAGT 890
QY      221 GIYGIYLyValaIIeGIuHIsleuAlaLySGIuGIYAsnArgPheHIsPheAspIIleLyS 240
        |||
DB      891 GGTTGGAAAGCCATTAACATTGGCCAAACAGGAATAGATTTCATTTTGACATCAAA 950
QY      241 ProProLeuHIsHIsAlaLySAsnCySaspPheSerPheThrGIYLeuGIuHIsValIThr 260
        |||
DB      951 CTTCCCTTGATCATGCTAATAAATGTGATTTTCTTTTACTGACATTCAACAGCTTACT 1010
```

```
QY      261 AspLySIIleIleMetLySlySGIuLySGIuGIuGIYIIeGIuLySGIYGIuIIeLeuSer 280
        |||
DB      1011 GATTAATAATATATGAAAAAGAAAAAGAGAAAGTATTAAGAAAGGGCAATCTGTCT 1070
QY      281 SerAlaAlaAspIIleAlaIaIaThrValaIIeHIsThrMetAlaCySHIsleuValIleArg 300
        |||
DB      1071 TCAGCAGCAGACATTTGCTGCACAGTACACACACAAATGGCATTCATCTTGTGAAAAAGA 1130
QY      301 ThrHIsArgAlaIIeLeuPheCySlySGIuArgAspLeuLeuProGIuAsnAsnAlaVal 320
        |||
DB      1131 ACAATCGGCTATCTGTTTGTGTTGACAGAGACATTTGTTACTTCAAAATAATGCACTA 1190
QY      321 LeuValaIaSerGIYGIYValaIaIaSerAsnPhenIleArgArgAlaLeuGIuIIeLeu 340
        |||
DB      1191 CTGTTGTCATCTGTGTGTGTGTCGCAAGTAACTTCATATTCGCAAGGCTCTGGAATTTTA 1250
QY      341 ThrAsnAlaThrGIuCySthrLeuLeuCySProProProArgLeuCySthrAspAsnGIY 360
        |||
DB      1251 ACAAAACCAACACAGTGCACCTTGTGTGTCTCTCCACAGACTATGCACTGATAATGCG 1310
QY      361 IleMetIIeAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 380
        |||
DB      1311 ATATATGATTCATGATGAATGGATTTGAAGAACTACGTGCTGTGGCATTTTACATGAC 1370
QY      381 IIeGIuGIYIIeArgIYArgIYProLyCySProLeuGIYValaAspIIeSerLySGIuVal 400
        |||
DB      1371 ATGACAGGCAATCCGCTATGAAACCAAAATGCTCTTGGAGTAAAGCATATCAAAAGAACTT 1430
QY      401 GIYGIuAlaSerIIeLySValIProGIuLeuLySmetGIuIIe 414
        |||
DB      1431 GGAGAGCTTCCATTAAGTACCAATTAATAATGAGATA 1472
```

RESULT 2

US-10-649-273-1
Sequence 1, Application US/10649273
Publication No. US20040043407A1
GENERAL INFORMATION:
APPLICANT: Bristol-Myers Squibb Company
TITLE OR INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1
FILE REFERENCE: D0073 CNT
CURRENT APPLICATION NUMBER: US/10/649, 273
PRIOR APPLICATION NUMBER: US 60/266, 518
PRIOR FILING DATE: 2003-08-27
PRIOR FILING DATE: 2001-02-05
PRIOR FILING DATE: 2002-02-05
PRIOR FILING DATE: 2002-02-05
PRIOR APPLICATION NUMBER: US 60/282, 814
PRIOR FILING DATE: 2001-04-10
NUMBER OF SEQ ID NOS: 71
SOFTWARE: PatentIn version 3.2
SEQ ID NO 1
LENGTH: 2197
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (231)..(1472)
US-10-649-273-1

Alignment Scores:

Pred. No.:	3,35e-247	Length:	2197
Score:	2125.00	Matches:	414
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	19	Gaps:	0

US-10-649-273-2 (1-414) x US-10-649-273-1 (1-2197)

```
QY      1 MetLeuIleLeuThrIlySthraIagIValIphePheLySProSeLySArgLySValIYr 20
        |||
DB      231 ATGCTAATCTTGACTAGACTGACAGAGATTTTAAACATCAAAAGAAAGTTAT 290
```

QY 21 GIUPELEUARGSERPHEANPHEHISPROGLYTHLEUPHELEUHSILYSIIIEVALLEU 40
DB 291 GAATTTTAAAGAGATTATTTTCACTCGAAGCACTATTTCTCATATAAATAGATATG 350
QY 41 GIIYIIIEGLIUTHSERCYSAAPSPTRHIAAIAIValaValaAspGIUTHRGIVASNAI 60
DB 351 GGAAATTTGAAACTAGTGTATGATACAGACGCTGCTGTGTGATGATAAATCGAATGTG 410
QY 61 LEUGLYGLUAIAlIehisSergIntHrgIuValIhISleuYstHrgIyIIeValIPro 80
DB 411 TTGGAGAGAAACATACATTCCTCAAGTCAAGTTCAATTTAAAAACAGTGGATTTCTCT 470
QY 81 PROAIAIAGIINGIINUHSIARGIUAANIIEGINARGIIIEVALGINGIUALALEUSER 100
DB 471 CCAGACAGCTCAACAGCTTCCAGAGAAATATTCAACGAATAGTACAAAGAAAGCTTTCT 510
QY 101 AIASeRGIVAlaserProserAspLeuserAIIeAlaIthThrIIelYsProGIYLeu 120
DB 531 GCCAGTGGAGTCTCTCAAGTACCTCTCAGCAATTGCAACTATCAATAAACAGAGACTT 590
QY 121 AIAleuserLeuGIYValGIyleuserPheSerLeuGIleuValGIyGInleuYsIys 140
DB 591 GCTTTAAGCTTGGAGTGGCTTATCATTTAGCTTACAGCTGTAGGACAGTTAAATAAG 650
QY 141 PROPHIEIleProIehisIsmetGIUAANIISAlaleuThrIIeArgLeuThrAsnIys 160
DB 651 CCATTCATTTCCATTCATATATGAGGCTCAATGCACTTACTATTTAGTTGACCAATPAA 710
QY 161 VALGIUPHEProPheleuValleuLeuIIeSergIyIhISCYaleuLeuAlaleuVal 180
DB 711 GTAGAAATTTCTTTTATTTTACTTTCTTTGATTTCTGGAGGTCACGTCTGTGTGCAATAGTT 770
QY 181 GINGIYValaserAspPheleuLeuGIYIysSerLeuAspIIeAlaIProGIYAspMet 200
DB 771 CAAGAGATTTCAGATTTTCTGCTTCTTGAAAGTCTTTGACATACACACAGGTGACATG 830
QY 201 LEUAspIysValAIARgArgLeuserLeuIIelYsHISProGIYUcYsSerThrmMetSer 220
DB 831 CTTGACAGAGTGGCAAGAGACTTTCTTTATTAATAACCCAGAGTCTCCACCAAGAGT 890
QY 221 GIYGIYIysAlaIIEGLIuhISleuAlaIysGINIYAsnArgPheHISpHeAspIIelYs 240
DB 891 GGTGGGAAAGCCATAGAACATTTGGCCAAACAGAAATAGATTTTCACTTTTGACATCAAA 950
QY 241 PROProleuHISHISAlaIysAsnCYsAspPheSerPheThrGIYleuGINHISValIThr 260
DB 951 CTTCCCTTGACATCATGCTMAAAATGTGATTTTCTTTTAACTGGACTTCAMACAGTTACT 1010
QY 261 ASPlySIIeIleMetIysIysGIUySGIUGIUGIYIIIEGLIUYSGIYGINIleuser 280
DB 1011 GATTAATAATATATGAAAGAAAGAAAGAAAGAGATTAAGAGAGGAGGCAATCTGCTCT 1070
QY 281 SERAlaIAspIIeAlaIaIthrValGINHISThrmEtAlaCYSHISleuValIysARG 300
DB 1071 TCAGACAGCAGATTTGCTCCACAGTACAGCACAGATGCAATGTCATCTTGGAATAAG 1130
QY 301 THRIISAGAlaIIEleuPheCYsIysGINARgAspLeuLeuProGINAsnAlaValI 320
DB 1131 ACACATCGGGCTATCTGTTTGTGAAGCAGAGACTTGTACTCTCAATAATATGACAGTA 1190
QY 321 LEUValAlaserGIYValAlaserAspPheYrIIeARGArgAlaleuGIUleu 340
DB 1191 CTGGTTCATCTGGAGTGTGCGAAGTAACTTCTATATCTCGACAGCTCTGGAAATTTTA 1250
QY 341 THRAsnAlaThrGINCYstHrleuLeuCYsProProArgLeuCYstHrAspANGIY 360
DB 1251 ACAAAACGACACAGTGCATTTGTGTGTCTCTCCACAGCTAGTACATATATATGCG 1310
QY 361 IIEHETIIeAlaIATPAsnGIYIIIEGIUAGIleuArgAlaGIYleuGIYIIeLeuHISAsp 380
DB 1311 ATTATGATTGCATGGAATGTGATTAAGACACTACGTCGTGGCGCAATTTTACATGAC 1370

QY 381 IIEGLIUGIYIIeArgIYRGUProIyCYsProLeuGIYValaAspIIeSergIuValI 400
DB 1371 ATAGAAAGCATTCGCTATATGAACCAAAATGTCTCTTGAGATGACATATCAAAAGAGTT 1430
QY 401 GIYGLUAlaserIIelYsValIProGINleuYsmetGIUle 414
DB 1431 GGAGAGCTTCATATAAATGATACCAATTAATAATGGAGATA 1472
RESULT 3
US-10-651-722-1
; Sequence 1, Application US/10651722
; Publication No. US20040048302A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1
; FILE REFERENCE: D0073 DIV
; CURRENT APPLICATION NUMBER: US/10/651,722
; CURRENT FILING DATE: 2003-08-29
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 2197
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (231)..(1472)
US-10-651-722-1
Alignment Scores:
Pred. No.: 3,35e-247 Length: 2197
Score: 2125.00 Matches: 414
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0
US-10-649-273-2 (1-414) x US-10-651-722-1 (1-2197)
QY 1 MetLeuIIeLeuThrIysThrAlaGIYValIhPhePheIysProSerIysArgIysValITyr 20
DB 231 ATGCTAATCTTGTGACTAAGACTGACAGAGTTTCTTTTAAACCATCAAAAGGAAATTTAT 290
QY 21 GIUPELEUARGSERPHEANPHEHISPROGLYTHLEUPHELEUHSILYSIIIEVALLEU 40
DB 291 GAATTTTAAAGAGATTATTTTCACTCGAAGCACTATTTCTCATATAAATAGATATG 350
QY 41 GIIYIIIEGLIUTHSERCYSAAPSPTRHIAAIAIValaValaAspGIUTHRGIVASNAI 60
DB 351 GGAAATTTGAAACTAGTGTATGATACAGACGCTGCTGTGTGATGATAAATCGAATGTG 410
QY 61 LEUGLYGLUAIAlIehisSergIntHrgIuValIhISleuYstHrgIyIIeValIPro 80
DB 411 TTGGAGAGAAACATACATTCCTCAAGTCAAGTTCAATTTAAAAACAGTGGATTTCTCT 470
QY 81 PROAIAIAGIINGIINUHSIARGIUAANIIEGINARGIIIEVALGINGIUALALEUSER 100
DB 471 CCAGACAGCTCAACAGCTTCCAGAGAAATATTCAACGAATAGTACAAAGAAAGCTTTCT 510
QY 101 AIASeRGIVAlaserProserAspLeuserAIIeAlaIthThrIIelYsProGIYLeu 120
DB 531 GCCAGTGGAGTCTCTCAAGTACCTCTCAGCAATTGCAACTATCAATAAACAGAGACTT 590
QY 121 AIAleuserLeuGIYValGIyleuserPheSerLeuGIleuValGIyGInleuYsIys 140
DB 591 GCTTTAAGCTTGGAGTGGCTTATCATTTAGCTTACAGCTGTAGGACAGTTAAATAAG 650

Qy	141	ProPhel1ePro11eh1sh1smetGlu1Ahi1A1euthr11e1ag1euthr1As1y1s	160
Db	651	CCATTTCATTCCCATTCATTCATATGAGAGCTCATGCACTTCTATTAAGTTGACCAATAAA	710
Qy	161	Val1Glu1hePheProPheLeuVal1Leu1e11e1serG1yG1y1h1sCy1sLeu1eua1A1eVal1	180
Db	711	G7AGAAATTTCTTTTAAAGTTCTTTAGTTCTTTGATTTCTGAGAGTCACTGTCTGTTGGCATTAAGTT	770
Qy	181	G1ng1yVal1Ser1Asp1heLeu1e1e1uG1y1y1sSer1e1u1Asp11e1A1ProG1y1Asp1e1t	200
Db	771	CAAGAGATTTCAGATTTTCTGCTCTTGAAAGCTTTTGACATATGACACCAAGGTGACATG	830
Qy	201	Leu1Asp1y1sVal1A1a1a1g1a1g1e1u1Ser1e1u11e1y1sh1sP1roG1u1Cy1sSer1Th1Me1Ser	220
Db	831	CTTGACAAAGGTGGCAGAGAGACTTTCTTTAATAAACATCCAGAGTCTCCACATAGT	890
Qy	221	G1yG1y1y1sVal1A1e1G1u1h1sLeu1A1y1sG1ng1y1As1a1gPhe1h1sPhe1Asp11e1y1s	240
Db	891	GGTGGAAAGCCATATGAAACATTTGGCCCAACAGAGAAATAGATTTCAATTTTGACATCAAA	950
Qy	241	ProP1roLeu1h1sh1sA1a1y1As1nCy1sAsp1Phe1SerPhe1ThrG1y1e1uG1nh1sVal1Thr	260
Db	951	CTCCCTTGATCATATGCTAAATTTGATTTTCTTTTACTGACATTTCAACAGCTTACT	1010
Qy	261	Asp1y1s1e11e1e1e1y1s1y1sG1u1y1sG1u1g1u1y11e1G1u1y1sG1y1G1n11e1u1Ser	280
Db	1011	GATAAATATTAATGAAGAAAGAAAGAGAGAGATATTTGAAAGGGGCAATCTCTCT	1070
Qy	281	Ser1A1a1A1Asp11e1A1a1Th1Val1G1nh1sTh1Me1A1a1Cy1sh1sLeu1A1y1sArg	300
Db	1071	TCAGAGAGACATTTGCTGCCACAGTACAGACACAAATGGCATATCTTTGTCAAAAAGA	1130
Qy	301	Thr1h1sArG1A1a11e1LeuPheCy1s1y1sG1n1Arg1sP1e1u1e1uProG1n1Asn1Asn1A1Val	320
Db	1131	ACACATCGGGCTATTTCTGTTTGTATAGACAGAGAGACTTTGATACCTCAAAAATAGCAGTA	1190
Qy	321	LeuVal1A1As1e1rG1y1G1yVal1A1As1e1rAsn1Phe1Ty11e1A1g1a1g1A1e1uG1u11e1u	340
Db	1191	CTGGTTGCATCTGGGTGGTGGCAGAAAGTAACTTATATCCGACAGAGCTCTTGAAATTTTA	1250
Qy	341	Thr1Asn1A1a1ThrG1nCy1sTh1Leu1e1uCy1sP1roP1roA1g1e1uCy1sTh1Asp1AsnG1y	360
Db	1251	ACAAACGACACACAGTGCATTTGTTGTCTCTCCACAGCTATGACATGATATATGAC	1310
Qy	361	11e1e1c11e1A1a1Tr1AsnG1y11e1G1u1Arg1e1u1Arg1A1g1y1e1uG1y11e1e1u1h1sAsp	380
Db	1311	ATTATGATTCATGATGGAATGATATTAAGACTACGTGCTGGCATTTTACATATAC	1370
Qy	381	11e1G1u1y11e1A1y1Ty1G1u1P1ro1y1Cy1sP1ro1e1uG1y1Val1Asp11e1Ser1y1sG1u1A1	400
Db	1371	ATGAGAGGCATCCGCTATGATACCAAAAGTCTCTTGAGATGATATCAAAAGAAAGTT	1430
Qy	401	G1yG1u1A1As1e1r11e1y1sVal1ProG1n1e1u1y1sMetG1u11e1	414
Db	1431	GGAGAGCTTCATTAAGTACCAACAATTAATAAATGAGATAT	1472
RESULT 4			
US-10-480-988-36			
Sequence 36, Application US/104808988			
Publication No. US20050069877A1			
GENERAL INFORMATION:			
APPLICANT: GANDHI, Ameena R.; KABLE, Amy E.;			
APPLICANT: SWARNAKAR, Anita; HARALIA, April J.A.;			
APPLICANT: TRAN, Bao; DUGAN, Brendan M.;			
APPLICANT: WARREN, Bridget A.; ISON, Craig H.;			
APPLICANT: HONCHERU, Cynthia D.; NGUYEN, Daniel B.;			
APPLICANT: LU, Dying Aina M.; LEE, Ernestine A.;			
APPLICANT: YUE, Henry; FORSYTHE, Ian V.;			
APPLICANT: BARROSO, Ines; RAMQUAR, Jeyalakmi;			
APPLICANT: GRIFFIN, Jennifer A.; Li, Joana X.;			
APPLICANT: YANG, Junning; THANHAVEU, Kavitha;			
APPLICANT: GIETZEN, Kimberly J.; DING, Li;			
APPLICANT: BAUGHN, Mariah R.; BOROWSKI, Mark L.;			

```

; APPLICANT: YAO, Monique G.; CHAMLA, Narinder K.;
; APPLICANT: MASON, Patricia M.; GURURAJAN, Rajagopal;
; APPLICANT: LEE, Sally; BECHA, Shanya D.;
; APPLICANT: LEE, Soo Yeun; TRAN, Uyen K.;
; APPLICANT: ELIOT, Vicki S.; LUO, Wen;
; APPLICANT: SPRAGUE, William W.; TANG, Y.
; APPLICANT: LU, Yan; ZEBARJADIAN, Yeganeh
TITLE OF INVENTION: PROTEIN MODIFICATION AND MAINTENANCE MOLECULES
FILE REFERENCE: PF-1040 USN
CURRENT FILING DATE: 2003-12-16
PRIOR APPLICATION NUMBER: PCT/US02/19360
PRIOR FILING DATE: 2002-06-18
PRIOR APPLICATION NUMBER: US 60/300,508
PRIOR FILING DATE: 2001-06-22
PRIOR APPLICATION NUMBER: US 60/303,445
PRIOR FILING DATE: 2001-07-06
PRIOR APPLICATION NUMBER: US 60/305,405
PRIOR FILING DATE: 2001-07-13
PRIOR APPLICATION NUMBER: US 60/311,442
PRIOR FILING DATE: 2001-08-09
PRIOR APPLICATION NUMBER: US 60/314,821
PRIOR FILING DATE: 2001-08-24
PRIOR APPLICATION NUMBER: US 60/315,992
PRIOR FILING DATE: 2001-08-29
PRIOR APPLICATION NUMBER: US 60/378,205
PRIOR FILING DATE: 2002-05-03
NUMBER OF SEQ ID NOS: 56
SOFTWARE: PERL Program
SEQ ID NO 36
LENGTH: 2572
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No: 7632424CB1
US-10-480-988-36

Alignment Scores:
Pred. No.: 4,3e-247 Length: 2572
Score: 2125.00 Matches: 414
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 22 Gaps: 0

US-10-649-273-2 (1-414) x US-10-480-988-36 (1-2572)
QY 1 MetLeuIleLeuThrLysThrAlaGlyValPhePheLysProSerLysbArgLysValIYr 20
Db 144 AGGCTAATCTTACACTAGACTGACCTGAGAGGTTTATTTTAAACCATCAAAAGGAAGTTTAT 203
QY 21 GlupheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisIysIleValLeu 40
Db 204 GAATTTTAAAGAAAGTTTAAATTTTATCTCTGGAACCTATTTCTTCAATAAAATAGTATTG 263
QY 41 GilylIegIuThrSerCyAspAspThrAlaAlaIaValAlaAspGluThrGlyAsnVal 60
Db 264 GGAATGAAACTAGTGTGATGATACAGACGCTGCTGTGATGATGAACCTGAAATGTG 323
QY 61 LeuGlyGluAlaIleHisSerGlnThrGlyValHisLeuLysThrGlyGlyIleValPro 80
Db 324 TTGGGGAAGCAATATCAATTCCTCAAACTGAAAGTTCATTAAACAAGGTGGGATTTCTCT 383
QY 81 ProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100
Db 384 CCAAGCAGCTCAACAGCTTCAACAGAGAAATATTCAACGAATATATCAAGAAGCTCTTTCT 443
QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeu 120
Db 444 GCCAGTGGAGTCTCTCCAAGTGAACCTCTCAGCAATTCGAACATCAATAAACAGGACTT 503
QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLysLys 140

```



```
Db      504 GCTTTAGCTGGAGTGGCTTATCATTTAGCTTACAGTGGTAGGACGTTTAAAAAG 563
Qy      141 Propheileproilehishmetglualahisaleuthrilleargleuthrasnly 160
Db      564 CCATTCAATCCCATTCATATAGAGGCTCATGCACTTACTATTAGTTGACCAATATA 623
Qy      161 ValglupheproheuleuValleuleuilleserglyvhiacyseuleuValleuVal 180
Db      624 GTAGAAATTTCTTTTATTCTTTTGTGATTTCTGGAGGTCACTGCTGTGGCATTTAGT 683
Qy      181 GlnGlyValSerAspPheuleuleuGlylySerleuaspillealaproGlyaspMet 200
Db      684 CAAGAGTTTCAGATTTCTGCTTCTTGAAAGCTTTTGACATGACCAAGGTGACATG 743
Qy      201 LeuAspLyValAlaArgArgLeuSerLeuilleserGlyvhiaproGlyCyseSerThmetSer 220
Db      744 CTTGACAAAGTGGCAAGAGACTTTCTTTATTAATAACATCCAGAGTCTCCACCATGAGT 803
Qy      221 GlyGlyLyValAlaileGluhiSleuAlalyGlnGlyAsnaArgPhehisPheaspille 240
Db      804 GGTGGGAAAGCCATAGAACATTTGGCCAAACAAGAAATAGATTTCAATTTTGACATCA 863
Qy      241 ProProleuHishisAlalySaenCySaapPheSerPheThrGlyLeuGlnhisValThr 260
Db      864 CTTCCCTTGATCATGCTTAAAAATGTGATTTCTTTTACTGAGCTTCAACACGTTACT 923
Qy      261 AspLySileilemetLySlysglyuglyuglyileGlylySglnileleuser 280
Db      924 GATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 983
Qy      281 SerAlaAlaSerPileAlaAlaThrValGlnhisThrMetAlaCyshisSleuVallySarg 300
Db      984 TCAGCAGCAGACATTTGCTGCCACAGTACAGCACACATGSCATGTCATCTTGAAAAAG 1043
Qy      301 ThrhisSargAlaileleuPheCySlysglnaArgAspLeuProGlnAsnaAlaVal 320
Db      1044 ACACATCGGGCTATTTCTGTTGTAAAGCAGACGACTTTTACTTCAATAATATGACGTA 1103
Qy      321 LeuValAlaSerGlyGlyValAlaSerAspPheThrilleargArgAlaileuGluileu 340
Db      1104 CTGGTGCATCTGGGGGTGCTCCAGATTAATCTTATATCCGACAGCTCTGGAATTTTA 1163
Qy      341 ThrAsnaAlaThrGlnCyseThrleuLeuCySProProArgLeuCySThrAspAnGly 360
Db      1164 ACAACGCAACACATGCACTTTGTTGTCTCTCCCAAGCTATGCACTGATATATGCG 1223
Qy      361 IleMetIleAlaTrpAsnGlyileGluArgLeuArgAlaGlyLeuGlyileleuHisasp 380
Db      1224 ATTATGATTCATGGAATGATGATTAAGACTACGTGCTGGCTTGGAATTTTACATGAC 1283
Qy      381 IleGluGlyIleArgGlyGlyProlyCySProleuGlyValAspIleSerLySglnVal 400
Db      1284 ATAGAGGCATCCGCTATGAAACCAAAATGCTCTTGAGTGAGACATATCAAAAGAGTT 1343
Qy      401 GlyGluAlaSerIleLySValProGlnleuLyMetGluile 414
Db      1344 GGAAGAGCTTCCATTAAGGTACCAATTAATAAATGAGATA 1385

RESULT 5
US-10-067-443-21
; Sequence 21, Application US/10067443
; Publication No. US20030082782A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED I
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
```

```
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 21
; LENGTH: 1387
; TYPE: DNA
; ORGANISM: homo sapiens
US-10-067-443-21

Alignment Scores:
Pred. No.: 2,57e-243 Length: 1387
Score: 2090.50 Matches: 412
Percent Similarity: 93.85% Conservative: 0
Best Local Similarity: 93.85% Mismatches: 2
Query Match: 98.38% Indels: 25
DB: 15 Gaps: 1

US-10-649-273-2 (1-414) x US-10-067-443-21 (1-1387)

Qy      1 MetleuileleuthrlystrAlaGlyValPhePheLySProSerLySArgLyValTyr 20
Db      24 ATGCTATCTTGACTAGACTGACGAGGTTTTTTTAAACATCAAAAGAAAGTTAT 83
Qy      21 GlupheleuArgSerPheAsnPhehisProGlyThrleuPheleuHislySilleValleu 40
Db      84 GAATTTTAAAGATTTAATTTTCACTCGAAACACTATTTCTTCAATAAATAGTATG 143
Qy      41 GlyIleGlyThrSerCySaapPheThrAlaAlaValValAspGlyThrGlyAsnVal 60
Db      144 GGAATGAAACCTAGTGTGATGATACAGCAGCTGCTGTGTGTGATGAATGAAATGTG 203
Qy      61 LeuGlyGlyAlaIleHisSerGlnThrGlyValHisSleuLySThrGlyIleValPro 80
Db      204 TTGGAGAAACCATATACATTTCCCAACTGAGTCTATTAATAACAGTGGATTTGCT 263
Qy      81 ProAlaAlaGlnGlnleuHisSargLyAsnIleGlnArgIleValGlnGlnAlaSer 100
Db      264 CCAGCAGCTACACAGCTTCCACAGAAATATTAACAGAAATAGTAACAAAGCTCTTCT 323
Qy      101 AlaSerGlyValaSerProSerAspLeuSerAlaIleAlaThrThrIleLySProGlyLeu 120
Db      324 GCCAGTGGAGATCTCTCCAGATGACCTTCCAGCAATGCCAATCAATTAACAGGACT 383
Qy      121 AlaLeuSerleuGlyValGlyLeuSerPheSerleuGlnleuValGlyGlnleuLyS 140
Db      384 GCTTTAAGCTGGAGGTGGCTTATCATTTAGCTTACAGCTGGTAGGACGTTAAAAAG 443
Qy      141 ProPheileproilehishmetglualahisaleuthrilleargleuthrasnly 160
Db      444 CCATTCAATCCCATTCATATAGAGGCTCATGCACTTACTATTAGTTGACCAATATA 503
Qy      161 ValglupheproheuleuValleuleuilleserglyvhiacyseuleuValleuVal 180
Db      504 GTAGAAATTTCTTTTATTCTTTTGTGATTTCTGGAGGTCACTGCTGTGGCATTTAGT 563
Qy      181 GlnGlyValSerAspPheuleuleuGlylySerleuaspillealaproGlyaspMet 200
Db      564 CAAGAGTTTCAGATTTCTGCTTCTTGAAAGCTTTTGACATGACCAAGGTGACATG 623
Qy      201 LeuAspLyValAlaArgArgLeuSerLeuilleserGlyvhiaproGlyCyseSerThmetSer 220
Db      624 CTTGACAAAGTGGCAAGAGACTTTCTTTATTAATAACATCCAGAGTCTCCACCATGAGT 683
Qy      221 GlyGlyLyValAlaileGluhiSleuAlalyGlnGlyAsnaArgPhehisPheaspille 240
Db      684 GGTGGGAAAGCCATAGAACATTTGGCCAAACAAGAAATAGATTTCAATTTTGACATCA 743
Qy      241 ProProleuHishisAlalySaenCySaapPheSerPheThrGlyLeuGlnhisValThr 260
Db      744 CTTCCCTTGATCATGCTTAAAAATGTGATTTCTTTTACTGAGCTTCAACACGTTACT 803
Qy      261 AspLySileilemetLySlysglyuglyuglyile----- 273
```

Db 804 GATAAATATATATGAAAAAGAGAGATATATTCTAATTAGTAAGTTGAA 863
Qy 274 -----GluIys 275
Db 864 CAGATAATATTTCTGATTTGCTCTAAATAAGCTGCTATTCTCGAGTATGAGAG 923
Qy 276 GtynlntleuSerSerAlaIasprlleaIaIaThrValGlnHsthrMetAlaCys 295
Db 924 GGGCAAAATCTGTCTTCAGCAGACATCTGTCACAGTACAGACCAATGGCATGT 983
Qy 296 HisleuValIysArgThrHisArgAlaIleuPheCysLysGlnArgAspLeuPro 315
Db 984 CATCTTGGAAAAAGAACACATGGGCTATTCTGTTTGTAAACAGAGACATTGTACCT 1043
Qy 316 GlnAsnAsnAlaValIleuValAlaSerGlyGlyValAlaSerAsnPheTyrIleArg 335
Db 1044 CAATAATATGCACTGCTGGTTCATCTGGTGTGCAAGTAACTTATATCCGACGA 1103
Qy 336 AlaIeuGluIleuThrAsnAlaThrGlnCysThrIleuLeuCysProProArgLeu 355
Db 1104 GCTCGAAATTTTAAACAAACGCAACAGTCACTTGTGTCTCTCCGACACTA 1163
Qy 356 CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeu 375
Db 1164 TGCACCTGATATGCGCATTAATGATGCAATGATGATGTAAGACTACGTGGCTTG 1223
Qy 376 GlyIleLeuHisAspIleGluGlyIleArgTyrGluProLysCysProLeuGlyValAsp 395
Db 1224 GGCACTTTATCATGACATTAAGGCAATCCGCTATGAACCAAAATGTCTCTTGGAGTAGAC 1283
Qy 396 IleSerLysGluValGlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 414
Db 1284 ATATCAAAAGAAAGTTGGAGAGCTTCATATAAAGTACCAATTAATAAATGAGATA 1340

RESULT 6
US-10-649-273-21
; Sequence 21, Application US/10649273
; Publication No. US20040043407A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1
; FILE REFERENCE: D0073 CNT
; CURRENT APPLICATION NUMBER: US/10/649, 273
; CURRENT FILING DATE: 2003-08-27
; PRIOR APPLICATION NUMBER: US 60/266, 518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067, 443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282, 814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 21
; LENGTH: 1387
; TYPE: DNA
; ORGANISM: homo sapiens
US-10-649-273-21

Alignment Scores:
Pred. No.: 2,57e-243 Length: 1387
Score: 2090.50 Matches: 412
Percent Similarity: 93.85% Conservative: 0
Best Local Similarity: 93.85% Mismatches: 2
Query Match: 98.38% Indels: 25
DB: 19 Gaps: 1

US-10-649-273-2 (1-414) x US-10-649-273-21 (1-1387)

Qy 1 MetLeuIleuPheThrIysThrAlaGlyValPhePheLysProSerLysArgLysValTyr 20
Db 24 ATCTAATCTTGTGATAAGCTGACAGAGTATTTTAAACCATCAAAAAGAAAGTTAT 83
Qy 21 GluPheLeuArgSerPheAsnPheHisProGlyThrIleuPheLysIleValIleu 40

Db 84 GATTTTAAAGAGTTTAAATTTTCTCCGAAACACTATTTCTTCAATAAATAGTATG 143
Qy 41 GlyIleGluThrSerCysAspSerThrAlaAlaIleValAspGluThrGlyAsnVal 60
Db 144 GGAATTTGAAACTAGTGTATGATGATACAGACGCTGTGTGTGATTAACACTGAAATGTG 203
Qy 61 LeuGlyGluAlaIleHisSerGlnThrGluValHisLeuLysThrGlyIleValPro 80
Db 204 TTGGAGAAAGCAATACATTCCTCAACTGAAAGTTCATTTAAAAACAGTGGATGTTCTT 263
Qy 81 ProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100
Db 264 CACAGAGCTCAACAGCTTACACAGAGAAATATTCACAGATATGACAGAAAGCTCTTCT 323
Qy 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeu 120
Db 324 GCCAGTGGAGTCTCTCAAGTGAACCTTCAGCAATGCAACTACCAATAAACCAGGACTT 383
Qy 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLys 140
Db 384 GCTTTAAGCTGGAGAGTGGCTTATCATTTTACCTTACAGCTGGTAGACAGTTAAAG 443
Qy 141 ProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLys 160
Db 444 CCATTCATTCCTCATTCATATGAGAGCTCATGCACTTACTATTAGTTGACCAATAAA 503
Qy 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180
Db 504 GTAGAAATTCCTTTTATGATCTTTGATTTCTGGAGAGTCACTGTCTGTGGCAATAGTT 563
Qy 181 GlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet 200
Db 564 CAAGAGTTTCAGATTTTCTGCTTCTTGAAAGCTTTGAGACTAGACACAGTGACATG 623
Qy 201 LeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGluCysSerThrMetSer 220
Db 624 CTTCAGACAGGTGGCAAAAGCTTTCTTATAATAAATCATCCAGAGTCTCCACATGAGT 683
Qy 221 GlyGlyLysAlaIleGluHisLeuAlaLysGlnLysAsnArgPheHisPheAspIleLys 240
Db 684 GGTGGGAAACCCATAGAGCAATTTGGCCAAACMAAGAAATGATTTCAATTTTGAACATAA 743
Qy 241 ProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260
Db 744 CTTCCCTTGATCATGCTAAATAATGTGATTTTCTTTACTGCACTTCAACAGCTTACT 803
Qy 261 AspLysIleIleMetLysGluLysGluGluGlyIle----- 273
Db 804 GATAAATATATGAAAAAGAGAGATATATTCTAATTAGTAAGTTGAA 863
Qy 274 -----GluIys 275
Db 864 CAGATAATATTTCTGATTTGCTCTAAATAAGCTGCTATTCTCGAGTATGAGAG 923
Qy 276 GtynlntleuSerSerAlaIasprlleaIaIaThrValGlnHsthrMetAlaCys 295
Db 924 GGGCAAAATCTGTCTTCAGCAGACATCTGTCACAGTACAGACCAATGGCATGT 983
Qy 296 HisleuValIysArgThrHisArgAlaIleuPheCysLysGlnArgAspLeuPro 315
Db 984 CATCTTGGAAAAAGAACACATGGGCTATTCTGTTTGTAAACAGAGACATTGTACCT 1043
Qy 316 GlnAsnAsnAlaValIleuValAlaSerGlyGlyValAlaSerAsnPheTyrIleArg 335
Db 1044 CAATAATATGCACTGCTGGTTCATCTGGTGTGCAAGTAACTTATATCCGACGA 1103
Qy 336 AlaIeuGluIleuThrAsnAlaThrGlnCysThrIleuLeuCysProProArgLeu 355
Db 1104 GCTCGAAATTTTAAACAAACGCAACAGTCACTTGTGTCTCTCCGACACTA 1163
Qy 356 CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeu 375


```

: NUMBER OF SEQ ID NOS: 1208
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 1047
: LENGTH: 1387
: TYPE: DNA
: ORGANISM: human
US-10-887-553A-1047

Alignment Scores:
Pred. No.: 2.57e-243      Length: 1387
Score: 2090.50           Matches: 412
Percent Similarity: 93.85%  Conservative: 0
Best Local Similarity: 93.85%  Mismatches: 2
Query Match: 98.38%      Indels: 25
DB: 22                  Gaps: 1

US-10-649-273-2 (1-414) x US-10-887-553A-1047 (1-1387)

QY 1 MetLeuIIeLeuThrLysThraGlyValIlePhePheLysProSerLysArgLysValTyr 20
Db 24 ATGTAATCTTGAAGCTGACAGAGAGTTTCTTAAACCATCAAGAAAGAAAGTTTAT 83
QY 21 GluPheLeuArgSerPheAsnDheHisProGlyThrLeuPheLeuHisLysIleValLeu 40
Db 84 GAATTTTAAAGAGTTTAAATTTTCACTCGAAACCTAATTTCTTCAATAAATAGTATTG 143
QY 41 GlyIleGluThrSerCysAspAspThrAlaIleAlaValValAspGluThrGlyAsnVal 60
Db 144 GGAATTTGAAACTAGTTGATGATACAGACAGCTGCTGTGTGATGATAAACTGGAAATGTG 203
QY 61 LeuGlyGluAlaIleHisSerGlnThrGluValHisLeuLysThrGlyIleValPro 80
Db 204 TTGGAGAGAGCAATACATTCCTCAAACTGAAGTTCATTTAAACCAAGGTGGATTGTCT 263
QY 81 ProAlaIleGluGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100
Db 264 CAGACAGCTCAACAGCTTACAGAGAAATATTCAACGAATTTACAGAGAGCTTTTCT 323
QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeu 120
Db 324 GCCAGAGAGCTCTCTCAAGTGAACCTCAGCAATGCAACTACCAATAAACAGAGACTT 383
QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLysVal 140
Db 384 GCTTTAAGCTGGAGAGTGGCTTATCATTTAGCTTACAGCTGGTAGAGCAGTTAAAAAG 443
QY 141 ProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLys 160
Db 444 CCAATTCATTCCTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 503
QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180
Db 504 GATGATTTCTCTTTTATGTTCTTTGATTTCTTGGAGGTCACTGCTGTGGCATTAAGT 563
QY 181 GlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet 200
Db 564 CAAGGAGTTTCAGATTTTCTGCTTCTTGGAAAGCTTTGACATGACCAAGGATGACATG 623
QY 201 LeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGluCysSerThrMetSer 220
Db 624 CTTCAGACAGGTGGCAAGAACATTTCTTTAATTAACATTCAGAGTCTCCACCATGAGT 683
QY 221 GlyGlyLysAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLys 240
Db 684 GGTGGAGAAACCTATAGCATTTGGCCAAACAGAAATATGATTTCAATTTTGAATCAAA 743
QY 241 ProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260
Db 744 CTTCCCTTGATCATGCTAAATAATTTGATTTTCTTTTCTGACCTTCAACAGCTTACT 803
QY 261 AspLysGluIleLeuLysLysGluLysGluGluGlyIle----- 273
Db 804 GATTAATAATATATGAAAAAGAAAAAGAGAGAGTATATTTCTTAATTAAGTAAAGTTGAA 863

```

```

QY 274 -----GluLys 275
Db 864 CAGATAAATATTCCTGATTTGTGCTTAAATAAGTCTCATTTCTGACAGTATGAGAG 923
QY 276 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 295
Db 924 GGGCAATCTGCTCTTCAAGAGAGCATTTGCTGCGCACAGTACAGACCAATATGGCATGT 983
QY 296 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLeuGlnArgAspLeuLeuPro 315
Db 984 CATCTGTGAAAGAACACATCGGCTATTCTTTTGTGAGAGAGAGCTTTGATACCT 1043
QY 316 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnDheThrIleArgArg 335
Db 1044 CAATAAATAGCAGTACTGGTGTGATCTGGTGGTGGCAAGTAACTTCTATATCCGACGA 1103
QY 336 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProAlaGluLeu 355
Db 1104 GCTCTGAAATTTTAAACAAACGACACAGTGCATTTGTGTCTCTCCACAGCTA 1163
QY 356 CysThrAspAsnGlyIleMetIleAlaTPAsnGlyIleGluArgLeuArgAlaGlyLeu 375
Db 1164 TGCACGTATATATGCAATATGATGATGATGATGATGATGATGATGATGATGATGATG 1223
QY 376 GlyIleLeuHisAspIleGluGlyIleArgTyrGluProLysCysProLeuGlyValAsp 395
Db 1224 GGCATTTTACATACATACATGAGGACATCCGCTATGAACCAAAAGTCTCTTGGAGTAC 1283
QY 396 IleSerLysGluValGlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 414
Db 1284 ATATCAAAAGAGTGGAGAGCTTCATTAAGTACCAATTAATAATGAGAGATA 1340

RESULT 9
US-10-012-140-6
: Sequence 6, Application US/10012140
: Publication No. US2003009017A1
: GENERAL INFORMATION:
: APPLICANT: Leibny, Kevin R.
: APPLICANT: Kapeller-Libermann, Rosana
: APPLICANT: Glucksmann, Maria A.
: TITLE OF INVENTION: 38650, 28472, 5495, 65507, 81588, AND
: TITLE OF INVENTION: 14354 METHODS AND COMPOSITIONS OF HUMAN PROTEINS AND USES
: TITLE OF INVENTION: THESEOF
: FILE REFERENCE: 381552004900
: CURRENT APPLICATION NUMBER: US/10/012,140
: PRIOR FILING DATE: 2001-11-08
: PRIOR APPLICATION NUMBER: 60/246,768
: PRIOR FILING DATE: 2000-11-08
: PRIOR APPLICATION NUMBER: 60/246,772
: PRIOR FILING DATE: 2000-11-08
: PRIOR APPLICATION NUMBER: 60/249,185
: PRIOR FILING DATE: 2000-11-15
: NUMBER OF SEQ ID NOS: 49
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 6
: LENGTH: 1245
: TYPE: DNA
: ORGANISM: Homo sapiens
US-10-012-140-6

Alignment Scores:
Pred. No.: 4.36e-243      Length: 1245
Score: 2088.00           Matches: 407
Percent Similarity: 99.03%  Conservative: 3
Best Local Similarity: 98.31%  Mismatches: 4
Query Match: 98.26%      Indels: 0
DB: 15                  Gaps: 0

US-10-649-273-2 (1-414) x US-10-012-140-6 (1-1245)

QY 1 MetLeuIIeLeuThrLysThraGlyValIlePhePheLysProSerLysArgLysValTyr 20

```

Db 1 ATGCTAATCTTGACCTAAGACCTGACAGAGTTTCTTTTAAACATCAAAAAGAAAGTTTAT 60
 Qy 21 GIUphELeuAArgSerPhehSnPhehISProGIyThrLeuPheLeuHISlySIIeValIleu 40
 Db 61 GAATTTTAAAGAAAGTTTAAATTTTCTCTGGAACACTATTTCTTCATATAAATAGATTTG 120
 Qy 41 GIYIIeGIuThrSerCySaSpaSpThraIaAlaIaValaValaAspGIuThrGIyaSnVal 60
 Db 121 GGAAATGGAACCTAGTTGTATGATACAGACGCTGCTGTGTATGATAAACTGGAATAATGTG 180
 Qy 61 LeuGIyGluAlaIleHISerGIInThrGIuValHISLeuLySerGIyGIYIIeValIPro 80
 Db 181 TTGGAGAGAGCAATACATTTCCAACTGAGATTCATTTAAAAACAGGTGGATTTCTCT 240
 Qy 81 ProAlaIaGIInGIInLeuHISaRgIuAenIIeGIInaRgIIeValGIInGIInaLeuSer 100
 Db 241 CCAGACAGCTCAACAGCTTCCACAGAGAAATATTCAACGAATAGTACAGAGAGCTTTTCT 300
 Qy 101 AlaSerGIyAlaSerProSerAspLeuSerAlaIleAlaThrThrIIeLyProGIyLeu 120
 Db 301 GCCAGTGAAGTCTCTCCAAAGTACCTCTCAGCAATTCAGCACTAAACCAAGAGACTT 360
 Qy 121 AlaLeuSerLeuGIYValGIYLeuSerPheSerLeuGIInLeuValGIYGIInLeuLySIV 140
 Db 361 GCTTTAAGCTTGAGAGTGGAGCTTATCATTTAGCTTACAGCTGGAGAGCAATTAAGAAAG 420
 Qy 141 ProPheIIeProIIeHISHISmetGIuAlaHISaIleuThrIIeArgLeuThraSnLyS 160
 Db 421 CCATTCATTTCCATTCATCATATGAGAGCTCATGCACTTACATTAAGTTGACCAATAAA 480
 Qy 161 ValGIuPheProPheLeuValIleuLeuIIeSerGIyGIYHISCySLeuLeuAlaLeuVal 180
 Db 481 GTAGAAATTTCTTTTGTAGTTCTTTGTGATTTCTGAGAGGTCACGTCTGTGGCATTAATT 540
 Qy 181 GIInGIYValISerAspPheLeuLeuLeuGIYLySserLeuAspIIeAlaProGIyAspMet 200
 Db 541 CAAGAGATTTCAGATTTCTGCTTTTGAAAGCTTTTGAGATACAGACAGAGTGCACAG 600
 Qy 201 LeuAspLySValaIaArgArgLeuSerLeuIIeLySHISProGIYCySerThrMetSer 220
 Db 601 CTTGACAAAGTGGCAGAGAGACTTTCTTAATTAACATCCAGAGTGCCTCCACATGAGT 660
 Qy 221 GIYGIYLySaIaIIeGIuHISLeuAlaLySGInGIYAsnArgPheHISpHeAspIIeLyS 240
 Db 661 GGTGGAAAGCCATAGAACATTTGGCCAAACAGAAATAGATTTTACATTTTACATCAAA 720
 Qy 241 ProProLeuHISHISaIeLySaSnCySaSpPheSerPheThrGIYLeuGIInHISValIThr 260
 Db 721 CCTCCCTTGCAATCATGTCTAAATGATGATTTTCTTTTACCTGCACTTCAACACGTTACT 780
 Qy 261 AspLySIIeIleMetLySLeuLyGIuLyGIuGIYIIeGIuLyGIYGIInIIeLeuSer 280
 Db 781 GATTAATAATTAAGAAACAGGAAACAGAAAGGATTAAGAAAGGCGCAAAATCCCTGCT 840
 Qy 281 SerAlaIaAspIIeAlaAlaThrValGIInHISThrMetAlaCySHISLeuValIySArg 300
 Db 841 TCAGCAGACAGACTTGTGCTGCCACAGTACAGACACAAATGCGATGTCATCTTGTGAAGA 900
 Qy 301 ThrHISaRgAlaIleLeuPheCySlySGInaRgaSpLeuLeuProGIInaSnAlaValI 320
 Db 901 ACAATCGGGCTATCTGTTTGTGTAAGCAGAGACTGTACTCAAAATGAATACAGTA 960
 Qy 321 LeuValaIaSerGIYGIYValaIaSerAspPheThrIIeArgArgAlaLeuGIInIIeLeu 340
 Db 961 CTGCTTCATCTGTGTGTGTGTGCAAGTAACTTAATATCCGACAGCTCTGGAATTTTAA 1020
 Qy 341 ThrAsnAlaThrGIInCySThrLeuLeuCySProProArgLeuCySThraSpAsnGIY 360
 Db 1021 ACAACCGCAACACAGTGCATTTGTGTGTGTCTCTCCAGACTATAGCATATATAGGC 1080
 Qy 361 IleMetIIeAlaATPAsnGIYIIeGIuArgLeuArgAlaGIYLeuGIYIIeLeuHISAsp 380
 Db 1081 ATATATGATTGATGGAATGATTAAGAAAGACTAGCTGTGCTGGCATTTTATCATGAC 1140

Qy 381 IIeGIuGIYIIeArgYIYrgIuProLyCySaProLeuGIYValaAspIIeSerLyGIuVal 400
 Db 1141 ATAGAAGCATCCGCTATGAAACCAAAATGTCTCTTGAGATGACATATCAAAAGAGTT 1200
 Qy 401 GIYGIuAlaSerIIeLySValProGIInLeuHISmetGIuIIe 414
 Db 1201 GGAAAGCTTCATTAAGATACCAAAATTAATAATGAGATA 1242

RESULT 10
 US-10-012-140-4
 ; Sequence 4, Application US/10012140
 ; Publication No. US2003009017A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Leiby, Kevin R.
 ; APPLICANT: Kapeller-Libermann, Rosana
 ; APPLICANT: Glucksmann, Maria A.
 ; TITLE OF INVENTION: 38650, 28472, 5495, 65507, 81586, AND
 ; TITLE OF INVENTION: 14354 METHODS AND COMPOSITIONS OF HUMAN PROTEINS AND USES
 ; FILE REFERENCE: 381552004900
 ; CURRENT APPLICATION NUMBER: US/10/012,140
 ; PRIOR FILING DATE: 2001-11-08
 ; PRIOR APPLICATION NUMBER: 60/246,768
 ; PRIOR FILING DATE: 2000-11-08
 ; PRIOR APPLICATION NUMBER: 60/246,772
 ; PRIOR FILING DATE: 2000-11-08
 ; PRIOR APPLICATION NUMBER: 60/249,185
 ; PRIOR FILING DATE: 2000-11-15
 ; NUMBER OF SEQ ID NOS: 49
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 4
 ; LENGTH: 1820
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (146)...(1390)
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (1)...(1820)
 ; OTHER INFORMATION: n = A,T,C or G
 ; US-10-012-140-4

Alignment Scores:
 Pred. No.: 7,95e-243 Length: 1820
 Score: 2088.00 Matches: 407
 Percent Similarity: 99.03% Conservative: 3
 Best Local Similarity: 98.31% Mismatches: 4
 Query Match: 98.26% Indels: 0
 DB: 15 Gaps: 0

US-10-649-273-2 (1-414) x US-10-012-140-4 (1-1820)
 Qy 1 MetLeuIIeLeuThrIySerThraIaGIYValPhePheLySProSerLySArgValIYr 20
 Db 146 ATGCTAATCTTGACCTAAGACCTGACAGAGTTTCTTTTAAACATCAAAAAGAAAGTTAT 205
 Qy 21 GIUphELeuAArgSerPhehSnPhehISProGIYThrLeuPheLeuHISlySIIeValIleu 40
 Db 206 GAATTTTAAAGAAAGTTTAAATTTTCACTCGAAACACTATTTCTTCATAAATAGATTTG 265
 Qy 41 GIYIIeGIuThrSerCySaSpaSpThraIaAlaIaValaValaAspGIuThrGIyaSnVal 60
 Db 266 GGAAATGGAACCTAGTTGTATGATATACAGACGCTGCTGTGTATGATAAACTGGAATAATGTG 325
 Qy 61 LeuGIYGIuAlaIleHISerGIInThrGIuValHISLeuLySerGIYGIYIIeValIPro 80
 Db 326 TTGGAGAGAGCAATACATTTCCAACTGAGATTCATTTAAAAACAGGTGGATTTCTCT 385
 Qy 81 ProAlaIaGIInGIInLeuHISaRgIuAenIIeGIInaRgIIeValGIInGIInaLeuSer 100
 Db 386 CCAGACAGCTCAACAGCTTCCACAGAGAAATATTCAACGAATAGTACAGAGAGCTCTTCT 445

```
QY 101 AlAserGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeu 120
Db 446 GCCAGTGAAGCTCTCCAAAGGACCTTCAGCAATTCGAACTACCAATAACACAGACTT 505
QY 121 AlaleuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLysLys 140
Db 506 GCTTTAAGCCTGGAGAGTGGCTTATCATTTAGCTTACAGCTGTGTGAGCAAGTAAAAAG 565
QY 141 ProPheIleProIleHisIleMetGlnAlaHisAlaLeuThrIleArgLeuThrAsnLys 160
Db 566 CCATTCATTCCTCCATCATATGAGAGGCTCATGCATCTACTATAGGTGACCAATATAA 625
QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180
Db 626 GTAAATATTCCTTTTATGATTCCTTTGATTCGTGAGGCTGCTGTGTGACATTAAGT 685
QY 181 GlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyLysMet 200
Db 686 CAAGGAGTTTCAGATTTCTGCTTTTGGAAAGCTTTTGACATACAGCAGGTGACATG 745
QY 201 LeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlyCysSerThrMetSer 220
Db 746 CTTCAGCAAGTGGCAAGAACTTTCTTTAATAAAACATCCAGAGTCTCCACCATAGT 805
QY 221 GlyGlyLysAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLys 240
Db 806 GGTGGAAAGCCATGAAACATTTGGCCAAACAAAGAAATGATTCATTTTGACATCAAA 865
QY 241 ProPheLeuHisIleAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260
Db 866 CCTCCCTTCATCATGCTTAAATATGATGATTTTCTTTTACCTGCACTTCAACGTTACT 925
QY 261 AspLysIleIleMetLysGlnLysGlnGlyGlyIleGlnLysGlyGlnIleLeuSer 280
Db 926 GATTAATAATATGAAACACGAAACAAAGGAAAGATTTGAGAGGGGCAAAATCTGTCT 985
QY 281 SerAlaAlaAspIleAlaAlaThrValGlnHisIleMetAlaCysHisLeuValLysArg 300
Db 986 TCAGCAGCAGACATGCTGCCACAGTACAGCACACATATGTCATCTTGTGAAAGAG 1045
QY 301 ThrHisArgAlaIleIleLeuPheCysLysGlnArgAspLeuProGlnAsnAlaVal 320
Db 1046 ACAACATCGGGCTATTCCTGTTGTTGTAAGCAGAGACTGTGTACTCAAAATTAATCGACTA 1105
QY 321 LeuValAlaSerGlyGlyValAlaSerAsnPheThrIleArgArgAlaLeuGlnIleLeu 340
Db 1106 CTGGTTGACTCTGGTGGTGTCCGAGTAACTTCTATATCCGAGAGCTCGGAAATTTTA 1165
QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeuCysThrAspAsnGly 360
Db 1166 ACAAAACGCAACACAGTGCATTTGTTGTGTCTCTCCAGACTATGCTGATTAATAGGC 1225
QY 361 IleMetIleAlaIleAsnGlyIleGlnArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380
Db 1226 ATTATGATTTGACATGAAATGATTAAGAAAGCTAGCTGTGGCATTTTACATGAC 1285
QY 381 IleGlnGlyIleArgGlyArgProLysCysPProLeuGlyValAspIleSerLysGlnVal 400
Db 1286 ATAGAAAGGATCCGCTATGAAACCAAAATGCTCTTGAGATGACATATCAAAAGAGTT 1345
QY 401 GlyGlnAlaSerIleLysValProGlnLeuLysMetGlnIle 414
Db 1346 GGAGAAAGCTTCCATTAAGTACCAACATTAATAATGAGATA 1387

RESULT 11
US-10-094-749-400
; Sequence 400, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
```

```
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YUKI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA
; FILE REFERENCE: 08435/0160
; CURRENT APPLICATION NUMBER: US/10/094, 749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350, 435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 400
; LENGTH: 2208
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-094-749-400

Alignment Scores:
Pred. No.: 3 65e-225 Length: 2208
Score: 1944.00 Matches: 386
Percent Similarity: 93.24% Conservative: 0
Best Local Similarity: 93.24% Mismatches: 4
Query Match: 91.48% Indels: 24
DB: 18 Gaps: 1

US-10-649-273-2 (1-414) x US-10-094-749-400 (1-2208)

QY 1 MetLeuIleLeuThrIleThrAlaGlyValPhePheLysProSerLysArgLysValIle 20
Db 344 ATGCTATCTTGACTTAAGCTGACGAGATTTTATTAACCATCAAAAGAAAGTTTAT 403
QY 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisIleValLeu 40
Db 404 GAATTTTAAAGAAAGTTTAAATTTTCATCTGGAACACTATTTCTTCATTAATAATGATTTG 463
QY 41 GlyIleGlnPheSerCysAspAspThrAlaAlaValValAspGluThrGlyAsnVal 60
Db 464 GGAATTAATACTATTTGATGATACAGACCTGCTGTGTGATGAACAGGAAATGCG 523
QY 61 LeuGlyGlnAlaIleHisSerGlnThrGlyValHisLeuLysThrGlyGlyIleValPro 80
Db 524 TTGGGAGAAAGCAATACATTCCTCAAACTGAATTCATTTAAAAACAGTGGGATTTCT 583
QY 81 ProAlaIleGlnGlnLeuHisArgGlnAsnIleGlnArgIleValGlnGlnAlaLeuSer 100
Db 584 CCACAGCTCAACAGCTTCAACAGAAATATTCACCAAAATGATCAAGAAAGCTCTTCT 643
QY 101 AlAserGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeu 120
Db 644 GCCAGTGAAGCTCTCCAAAGTGAACCTTCAGCAATTCGACATCACTAATAACACAGACTT 703
QY 121 AlaleuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLysLys 140
Db 704 GCTTTAAGCCTGGAGAGTGGCTTATCATTTAGCTTACAGCTGTGTGAGCAAGTAAAAAG 763
QY 141 ProPheIleProIleHisIleMetGlnAlaHisAlaLeuThrIleArgLeuThrAsnLys 160
Db 764 CCATTCATTCCTCCATCATATGAGAGGCTCATGCATCTACTATAGGTGACCAATATAA 823
```

QY 161 ValGluPheProPheLeuValLeuIleSerGlyYHisCysLeuLeuAlaLeuVal 180
 Db 824 GAGAGATTTCTTTTATGTTCTTTGATTTCTGGAGGTCACGTCTGTTGGCAATTAGT 883
 QY 181 GInGlyValSerAspPheLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet 200
 Db 884 CAAGAGATTTCAGATTTTCTGCTTCTTGGAAAGTCTTTGGACATAGCACAGGTGACATG 943
 QY 201 LeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlyCysSerThrMetSer 220
 Db 944 CTTCAGAGGTGGCAAGAGACTTCTTAAATTAACATCCAGAGTCTCACACATGAGT 1003
 QY 221 GlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLys 240
 Db 1004 GGTGGAAAGCCATAGAACATTGGCCAAACAGAAATAGATTTCATTTTGACATCAACAA 1063
 QY 241 ProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260
 Db 1064 CCTCCCTTGATCATAGCTTAATAATTTGATTTCTTTTACTGGACTTCACACACGTTACT 1123
 QY 261 AspLysIleIleMetLysLysGlnLysGlnGlyIleGluLysGlnIleLeuSer 280
 Db 1124 GATTAATAATATATGAAAAAGAAAAAGAAAGATTTGAGAGAGGGCAATCTGTCT 1183
 QY 281 SerAlaAlaAspIleAlaAlaThrValGlnHisPheMetAlaCysHisLeuValLysArg 300
 Db 1184 TCAGACAGCAGACATTGCTCCACAGTACAGCACACAAATGTCATCTTGGAAGAAAG 1243
 QY 301 ThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuProGlnAsnAlaVal 320
 Db 1244 ACACATCGGGCTATTCCTTTGTTGAAGCAGAGAGACTTGTATCTTAATAATATATGACGA 1303
 QY 321 LeuValAlaSerGlyValAlaLysAsnPheTyrIleArgArgAlaLeuGlnIleLeu 340
 Db 1304 CTGCTTGATCTGGTGGTCTCCAGATTAATCTTGATCTCCAGAGCTTCGAAATTTT 1363
 QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeuCysThrAspGln 360
 Db 1364 ACAACGCAACACAGTGCATTTGTTGTCTCTCCAGACTAGTATGACATGATATGAG 1423
 QY 361 IleMetIleAlaThrAsnGlyIleGluArgLeuAlaGlyLeuGlyIleLeuHisAsp 380
 Db 1424 ATTAATGATTGCA-----TGATGTCTCTTGAGTACATATCAAAAGAAAT 1435
 QY 381 IleGluGlyIleArgTyrGluProLysCysProLeuGlyValAspIleSerLysGlnVal 400
 Db 1436 -----TGATGTCTCTTGAGTACATATCAAAAGAAAT 1471
 QY 401 GlyGluAlaSerIleLysValProGlnLeuLysMetGlnIle 414
 Db 1472 GGAGAGCTTCATTAAGTACACATTAATAATGAGATA 1513
 RESULT 12
 US-10-723-860-7447
 ; Sequence 7447, Application US/10723860
 ; Publication No. US20040253606A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Aziz, Natsaba
 ; APPLICANT: Ginsburg, Wendy M.
 ; APPLICANT: Zlotnik, Albert
 ; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
 ; FILE REFERENCE: 05882, 0193, NUS01
 ; CURRENT APPLICATION NUMBER: US/10/723,860
 ; PRIOR FILING DATE: 2003-11-26
 ; PRIOR APPLICATION NUMBER: 60/429,739
 ; NUMBER OF SEQ ID NOS: 11-26
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO: 7447
 ; LENGTH: 2890
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens

; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (646)..(657)
 ; OTHER INFORMATION: n is a, c, g, or t
 US-10-723-860-7447
 Alignment Scores:
 Pred. No.: 5.59e-225 Length: 2890
 Score: 1944.00 Matches: 386
 Percent Similarity: 93.24% Conservative: 0
 Best Local Similarity: 93.24% Mismatches: 4
 Query Match: 91.48% Indels: 24
 DB: 21 Gaps: 1
 US-10-649-273-2 (1-414) x US-10-723-860-7447 (1-2890)
 QY 1 MetLeuIleLeuThrIleThrAlaGlyValPhePheLysProSerLysArgLysValIle 20
 Db 1001 ATGCTAATCTTGACTTAAGACTGCGAGGATTTTAAACCATCAAAAGAAAGTTTAT 1060
 QY 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisLysIleValLeu 40
 Db 1061 GAATTTTAAGAAGTTTAAATTTTCATCTGGAACACTAATTTCTTAATAATATGATATG 1120
 QY 41 GlyIleGluThrSerCysAspAspThrAlaAlaAlaValAlaAspGluThrGlyAsnVal 60
 Db 1121 GGAATTTGAACCTAGTTGTGATGATACAGCAGCTGCTGTGGTGAATCTGAAATGCTG 1180
 QY 61 LeuGlyGluAlaIleHisSerGlnThrGlnValHisLeuLysThrGlyIleValPro 80
 Db 1181 TTGGAGAAACAAATACATCTCCAAACGAAATTCATTTAAACAGGTGGAGATTTGCT 1240
 QY 81 ProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGlnAlaLeuSer 100
 Db 1241 CCGACAGCTCAACAGCTTCCAGAGAAATATTCAGAAATAGTACAAAGAGCTTTTCT 1300
 QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeu 120
 Db 1301 GCCAGTGAAGTCTCTCAATGACCTTCAGCAATTCGCAATCAATCAATCAATCAATCA 1360
 QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLys 140
 Db 1361 GCTTTAAGCTCGGAGAGGCTTATCATTTAGCTTACAGCTGAGTACAGTTAAAAAG 1420
 QY 141 ProPheIleProIleHisHisMetGlnAlaHisAlaLeuThrIleArgLeuThrAsnLys 160
 Db 1421 CCAATTAATCCCATTAATATATGAGGCTCAGTCACTTAATAGTTGACCAATAA 1480
 QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyYHisCysLeuLeuAlaLeuVal 180
 Db 1481 GTAGATTTCTTTTATGTTCTTTGATTTCTGAGGTACATGCTGTGTGGCATTTAGTT 1540
 QY 181 GInGlyValSerAspPheLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet 200
 Db 1541 CAAGAGTTTCAGATTTTCTGCTTCTTGAAGATCTTTGGACATAGACACAGGTGACATG 1600
 QY 201 LeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlyCysSerThrMetSer 220
 Db 1601 CTTCAGAGGTGGCAAGAGACTTCTTAAATTAACATCCAGAGTCTCCACATGAGT 1660
 QY 221 GlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLys 240
 Db 1661 GGTGGAAAGCCATAGAACATTGGCCAAACAGAAATAGATTTCATTTTGACATCAACAA 1720
 QY 241 ProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260
 Db 1721 CCTCCCTTGATCATAGCTTAATAATTTGATTTCTTTTACTGACATTCACACGTTACT 1780
 QY 261 AspLysIleIleMetLysLysGlnLysGlnGlyIleGluLysGlnIleLeuSer 280
 Db 1781 GATTAATAATATATGAAAAAGAAAAAGAAAGATTTGAGAAAGGGCAATCTGTCT 1840
 QY 281 SerAlaAlaAspIleAlaAlaThrValGlnHisPheMetAlaCysHisLeuValLysArg 300

Db 1841 TCAGCAGACAGATTGCTGCCACACTACAGCACAAATGCGATGTCATCTTGAAAACA 1900
Qy 301 ThrHisArgAlaIleLeuPheCysIysGlnArgAspLeuProGlnAsnAsnAlaVal 320
Db 1901 ACACATCGGCGCTATTCTCTTTTGTAAAGCAGAGACTTGTAACCTCAAAATATTCAGATA 1960
Qy 321 LeuValAlaSerGlyGlyValAlaSerAsnPheTyrIleArgArgAlaLeuGlnIleLeu 340
Db 1961 CTGGTTGCATCTGGGAGTGTGCGACAGTAACCTCTATCCGACAGCTCTGGAATTTTAA 2020
Qy 341 ThrAsnAlaThrGlnCysThrIleuLeuCysProProArgLeuCysThrAspAsnGly 360
Db 2021 ACAACCGACACACAGTGCATTTGTGTCTCTCCAGACTAGCATGATAATGAGC 2080
Qy 361 IleMetIleAlaTrpAsnGlyIleGlnArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380
Db 2081 ATATGATTGCA----- 2092
Qy 381 IleGlnGlyIleArgTyrGlnProIysCysProLeuGlyValAspIleSerIysGlyVal 400
Db 2093 -----TGATGTCTCTTGAGTAGACATATCAAAAGAAATT 2128
Qy 401 GlyGluAlaSerIleIysValProGlnLeuIysMetGlnIle 414
Db 2129 GGAGAAAGCTTCATAAAGTACCAATTAAATGAGATA 2170

RESULT 13

US-10-120-988-177
; Sequence 177, Application US/10120988
; Publication No. US20030219745A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Goodrich, Ryle
; APPLICANT: Liu, Chenghua
; APPLICANT: Ren, Feiyun
; APPLICANT: Wang, Dunrui
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US20030219745A1el Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 802CON
; CURRENT APPLICATION NUMBER: US/10/120, 988
; CURRENT FILING DATE: 2002-04-11
; PRIOR APPLICATION NUMBER: 09/774, 528
; PRIOR FILING DATE: 2001-01-30
; NUMBER OF SEQ ID NOS: 441
; SOFTWARE: pc_fl_genes Version 2.0
; SEQ ID NO 177
; LENGTH: 1416
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (205)..(1305)
US-10-120-988-177

Alignment Scores:

Pred. No.: 1,74e-201 Length: 1416
Score: 1747.00 Matches: 340
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.71% Mismatches: 0
Query Match: 82.21% Indels: 0
DB: 18 Gaps: 0

US-10-649-273-2 (1-414) x US-10-120-988-177 (1-1416)

Qy 74 LysThrGlyGlyIleValProProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArg 93
Db 280 AGAACGAGTGGATGTCTCTCCAGAGCTCAACAGCTTCAAGAAATATTCACAGCA 339
Qy 94 IleValGlnGlnAlaLeuSerAlaSerGlyValSerProSerAspLeuSerAlaIleAla 113
Db 340 ATAGTCAAGAAAGCTTTCTGCGAGTGAAGTCTCTCAAGTGAACCTCTCAGCAATTGCA 399

Qy 114 ThrThrIleLeuProGlyLeuAlaLeuSerIleGlyValGlyLeuSerPheSerIleGln 133
Db 400 ACTACATTAACCAAGGACTTCTTTAAAGCTGGAGGTGGGCTATCATTTAGCTTACAG 459
Qy 134 LeuValGlyGlnLeuIysAspPheIleProIleHisMetGlnAlaHisAlaLeu 153
Db 460 CTGGTAGACAGTAAAGCAATTCATTCATCATATGAGAGCTCATGACATT 519
Qy 154 ThrIleArgLeuThrAsnIysValGlnPheProPheLeuValLeuLeuIleSerGly 173
Db 520 ACTATTAGGTGACCAATTAAGTAGAATTTCTTTTATGTTCTTTGATTCTGGAGGT 579
Qy 174 HisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyIysSerLeu 193
Db 580 CACTGCTGTGGCATTAAGTTCAGAGGTTTCAGATTCTTCTCTTGGAAAGTCTTTG 639
Qy 194 AspIleAlaProGlyAspMetLeuAspIysValAlaArgArgLeuSerIleIysHis 213
Db 640 GACATAGCACACAGGTGACATGCTTGAACAAGTGGCAAGAAAGCTTCTTAAATTAACAT 699
Qy 214 ProGlnCysSerThrMetSerGlyGlyValAlaIleGlnHisLeuAlaIysGlnIysAsn 233
Db 700 CCAGAGTGTCCACCATGAGTGTGGAAAGCCATGAAACATTTGGCCAAACAGAAAT 759
Qy 234 ArgPheHisPheAspIleIysProProLeuHisHisAlaIysAsnCysAspPheSerPhe 253
Db 760 AGATTTCATTTTGAACATCAAACTCTCTTCATCATCTTAAATTTGATTTTCTTT 819
Qy 254 ThrGlyLeuGlnHisValThrAspIysIleIleMetIysGlnGlyGlnGlyIle 273
Db 820 ACTGCACTTCAACACGTTACTGATTAATAATATAGAAAGAAAGAAAGGATATT 879
Qy 274 GlnIysGlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMet 293
Db 880 GAGAAAGGCGAAATCTGCTTCAGACAGACATTCCTCCACATGACACACACATG 939
Qy 294 AlaCysHisLeuValIysArgThrHisArgAlaIleLeuPheCysIysGlnArgAspLeu 313
Db 940 GCATGTCATCTTGGAAGAAACACATCGGGCTATCTGTTTGTAAAGCAGAGACTTG 999
Qy 314 LeuProGlnAsnAsnAlaValLeuValAlaSerGlyValAlaAspAsnPheTyrIle 333
Db 1000 TTACTCTAAATTAATGACAGTACGTTGTCATCTGTGTGTGCGAAGTAATCTTATATC 1059
Qy 334 ArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProPro 353
Db 1060 CGCAGAGCTCTGGAAATTTTAAACAACGCAACAGTGCACCTTTGTGTCTCTCC 1119
Qy 354 ArgLeuCysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGlnArgLeuArgAla 373
Db 1120 AGACTATGACCTGTAATGATGATTAATGATGATGATGATGATGATGATGATGATGAT 1179
Qy 374 GlyLeuGlyIleLeuHisAspIleGlyIleArgTyrGlnProIysCysProLeuGly 393
Db 1180 GGTGTGGCATTTTACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1239
Qy 394 ValAspIleSerIysGlnValGlyGlnAlaSerIleIysValProGlnLeuIysMetGln 413
Db 1240 GTAGACATATCAAAAGAAAGTGGAGAGCTTCCATTAAGTACCAATTAATAATGAG 1299
Qy 414 Ile 414
Db 1300 ATA 1302

RESULT 14

US-10-067-443-23
; Sequence 23, Application US/10067443
; Publication No. US20030082782A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED
; TITLE OF INVENTION: SPINDAL CORD, ME-1


```
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO: 23
; LENGTH: 1526
; TYPE: DNA
; ORGANISM: homo sapiens
;
US-10-067-443-23

Alignment Scores:
Pred. No.: 2,29e-157 Length: 1526
Score: 1385.00 Matches: 267
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 65.18% Indels: 0
Gaps: 0

US-10-649-273-2 (1-414) x US-10-067-443-23 (1-1526)

QY 148 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnIleValGluPheProPheLeuVal 167
Db 1 ATGAGAGCTCATGCACTTACTATTAGTTGACCAATAAGTGAATTCCTTTTATGTT 60

QY 168 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 187
Db 61 CTTTGTGATTTCTTGAGAGCTGCTGTTGGCATTAGTTCAGAGGATTTCAGATTTTCTG 120

QY 188 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 207
Db 121 CTTCTTGAAGAGCTTTTGACATAGCACAGGTGACATGCTTGACAGGTGGCAAGAGA 180

QY 208 LeuSerLeuIleLysHisProGluCysSerThrMetSerGlyGlyValAlaIleGlnHis 227
Db 181 CTTTCTTTAATAAACAATCCAGAGTCTCCACCATGATGGTGGGAAAGCCATAGAACT 240

QY 228 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 247
Db 241 TTGGCCCAACAGAGAAATAGATTTCATTTTGAACATCAAACTCCCTTCATCATCTAA 300

QY 248 AsnCysAspPheSerPheThrGlyLeuGlnHisValIleThrAspLysIleIleMetLysLys 267
Db 301 AATTGTGATTTTCTTTTACTGACCTTCAACACGTTACTGATAAATATATGAAAAAG 360

QY 268 GluLysGlnGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaIle 287
Db 361 GAAAAAAGAGAGATATTGAGAAAGGGCAAACTCTGTTTCAGCAGCAGCATTCCTGCC 420

QY 288 ThrValGlnHisThrMetAlaCysHisLeuValIleLysArgThrHisArgAlaIleLeuPhe 307
Db 421 ACAGTACAGACACAAATGCGATGTCATCTTGTAAGAAACACACATCGGGCTATTCTGTT 480

QY 308 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyVal 327
Db 481 TGTAAAGAGAGAGCTGTGTAACCTCAAAATATGACATGCTGTCGTGTGCTGTC 540

QY 328 AlaSerAsnPheThrIleArgArgAlaLeuGlnIleLeuThrAsnAlaIleThrGlnCysThr 347
Db 541 GCAGATTAATCTTATATCCGAGAGCTCTGAAATTTTAAACAAAGCAACAGTGCCT 600

QY 348 LeuLeuCysProProArgLeuCysThrAspAsnGlyIleMetIleAlaIleArgAsnGly 367
Db 601 TTGTGTGCTCTCTCCCAAGCTATGCACTGATATAGGCATTATGATTCATGCAATGCT 660

QY 368 TleGlnLysGlnGlyValGlyValGlyValIleLeuHisAspIleGlnGlyIleArgGlyGln 387
Db 661 ATGGAAGAGCTACGCTGCTGGCATTTATACATGACATAGAAGGATCGGCTATGTA 720
```

```
QY 388 ProLysCysProLeuGlyValAlaAspIleSerLysGlnValGlyAlaSerIleLysVal 407
Db 721 CCAAAATGCTCTTGGAGTAGACATATCAAAAGAAAGTTGGAGAGCTTCATATAAGTA 780

QY 408 ProGlnLeuLysMetGlnIle 414
Db 781 CCACAATTAAAAATGGAGATA 801

RESULT 15
US-10-649-273-23
; Sequence 23. Application US/10649273
; Publication No. US20040043407A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1
; FILE REFERENCE: D0073 CNT
; CURRENT APPLICATION NUMBER: US/10/649,273
; CURRENT FILING DATE: 2003-08-27
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO: 23
; LENGTH: 1526
; TYPE: DNA
; ORGANISM: homo sapiens
;
US-10-649-273-23

Alignment Scores:
Pred. No.: 2,29e-157 Length: 1526
Score: 1385.00 Matches: 267
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 65.18% Indels: 0
Gaps: 0

US-10-649-273-2 (1-414) x US-10-649-273-23 (1-1526)

QY 148 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnIleValGluPheProPheLeuVal 167
Db 1 ATGAGAGCTCATGCACTTACTATTAGTTGACCAATAAGTGAATTCCTTTTATGTT 60

QY 168 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 187
Db 61 CTTTGTGATTTCTTGAGAGCTGCTGTTGGCATTAGTTCAGAGGATTTCAGATTTTCTG 120

QY 188 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 207
Db 121 CTTTGTGAAAGCTTTTGACATAGCACAGGTGACATGCTTGACAGGTGGCAAGAGA 180

QY 208 LeuSerLeuIleLysHisProGluCysSerThrMetSerGlyGlyValAlaIleGlnHis 227
Db 181 CTTTCTTTAATAAACAATCCAGAGTCTCCACCATGATGGTGGGAAAGCCATAGAACT 240

QY 228 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 247
Db 241 TTGGCCCAACAGAGAAATAGATTTCATTTTGAACATCAAACTCCCTTCATCATCTAA 300

QY 248 AsnCysAspPheSerPheThrGlyLeuGlnHisValIleThrAspLysIleIleMetLysLys 267
Db 301 AATTGTGATTTTCTTTTACTGACCTTCAACACGTTACTGATAAATATATGAAAAAG 360

QY 268 GluLysGlnGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaIle 287
Db 361 GAAAAAAGAGAGATATTGAGAAAGGGCAAACTCTGTTTCAGCAGCAGCATTCCTGCC 420

QY 288 ThrValGlnHisThrMetAlaCysHisLeuValIleLysArgThrHisArgAlaIleLeuPhe 307
Db 421 ACAGTACAGACACAAATGCGATGTCATCTTGTAAGAAACACACATCGGGCTATTCTGTT 480
```

```

Qy 308 CysIysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyVal 327
Db 481 TGTAAAGCAGAGAGACTTGTACTCAAAATAATGACGACTGTTGCATCTGGTGTGTC 540
Qy 328 AlaSerAsnPhenylIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThr 347
Db 541 GCAAGTAACCTTATATCCGAGAGCTTGAAATTTTAAACAAACGCAACACAGTGCAC 600
Qy 348 LeuLeuCysProProArgLeuCysThrAspAsnGlyIleMetIleAlaTrpAsnGly 367
Db 601 TTGTTGTCTCTCTCCAGACTATGCACTGATATGCAATTATGCAATGCAATGCT 660
Qy 368 IleGlnArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlnGlyIleArgTyrglu 387
Db 661 ATTGAAGACTACGCTGCTGGCTTGCGCATTTTACATGACATAGAGGCATCCGCTATGAA 720
Qy 388 ProIysCysProLeuGlyValAspIleSerIysGlnValGlyGlnAlaSerIleIysVal 407
Db 721 CCAAAATGTCTCTTGAGTAGACATATCAAAAGAGTTGAGAGCTTCCATPAAAGTA 780
Qy 408 ProGlnLeuIysMetGluIle 414
Db 781 CCACATTTAAATGAGAGTA 801

```

Search completed: November 11, 2005, 03:16:23
 Job time : 1401.75 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: November 10, 2005, 16:52:13 ; Search time 8167.05 Seconds
(without alignments)
2456.267 Million cell updates/sec

Title: US-10-649-273-2

Perfect score: 2125 1 MLITKTAGVFKPSKRRVY.....DISKEVGASIKVPLKMEI 414

Sequence: 1 MLITKTAGVFKPSKRRVY.....DISKEVGASIKVPLKMEI 414

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Delop 6.0, Delext 7.0

Searched: 4708233 seqs, 2422767955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+ p2n model -DEV=xlp
-Q=/cgr21/USFPO.spool/p/US10649273/runat_02112005_091338_15550/app_query.fasta_1.1429
-DB=GenBdb1 -QFMT=faetap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=pcp -THR MAX=100 -THR MTN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pic -NORM=exc -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USRR=US10649273 @CGN 1 1 7742 @runat_02112005_091338_15550 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOC
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenBdb1:*
1: gb_ba:*
2: gb_ncg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2125	100.0	1908	9	BC011904 Homo sapi
2	2125	100.0	1907	6	AR428803 Sequence
3	2090.5	98.4	1387	6	AR428808 Sequence
4	2090.5	98.4	1387	9	HSA295148 Homo sapi

5	2088	98.3	1245	6	AX664697	AX664697 Sequence
6	2088	98.3	1820	6	AX664695	AX664695 Sequence
7	1944	91.5	2208	6	AX713716	AX713716 Sequence
8	1944	91.5	2208	9	AK055441	AK055441 Homo sapi
9	1835	86.4	1844	10	BC058172	BC058172 Sequence
10	1747	82.2	1416	6	AR541928	AR541928 Sequence
11	1725	81.2	1546	10	BC078974	BC078974 Rattus no
12	1385	65.2	1526	6	AR428809	AR428809 Sequence
13	1362	64.1	1522	6	AX934991	AX934991 Gallus ga
14	1273	59.9	1027	10	BC038910	BC038910 Mus muscu
15	1208	56.8	1558	9	BC030963	BC030963 Gallus ga
16	1193.5	56.2	8415	9	AC013468	AC013468 Homo sapi
17	1186.5	55.8	14364	6	AR428807	AR428807 Sequence
18	995.5	46.8	249601	2	AC114153	AC114153 Rattus no
19	995.5	46.8	308652	2	AC121478	AC121478 Rattus no
20	970	45.6	256751	10	AC122925	AC122925 Mus muscu
21	950.5	44.7	1109	10	BC051211	BC051211 Mus muscu
22	938	44.1	860	5	BX930694	BX930694 Gallus ga
23	827	38.9	875	6	CQ721898	CQ721898 Sequence
24	751.5	35.4	121251	5	AL591593	AL591593 Zebrafish
25	714.5	33.6	1576	3	AY051882	AY051882 Drosophila
26	714.5	33.6	1601	6	CQ606432	CQ606432 Sequence
27	710	33.4	1474	3	AK113378	AK113378 Ciona int
28	709	33.4	117322	5	AL672217	AL672217 Zebrafish
29	700.5	33.0	1385	6	BD157102	BD157102 Primer fo
30	700.5	33.0	1385	6	AX878239	AX878239 Sequence
31	700.5	33.0	1385	6	AX878236	AX878236 Homo sapi
32	687.5	32.4	3656	6	CQ606431	CQ606431 Sequence
33	687.5	32.4	14679	2	AC018262	AC018262 Drosophila
34	687.5	32.4	180263	3	AC010671	AC010671 Drosophila
35	687.5	32.4	207432	3	AE003513	AE003513 Drosophila
36	668.5	31.5	1443	8	AY024338	AY024338 Arabidops
37	668.5	31.5	1474	8	AY117283	AY117283 Arabidops
38	668.5	31.5	1567	8	AY063864	AY063864 Arabidops
39	662.5	31.2	1557	8	AY084577	AY084577 Arabidops
40	656	29.6	571	6	AK070912	AK070912 Oryza sat
41	629	29.6	571	6	BD154975	BD154975 Primer fo
42	629	29.6	571	6	AX874913	AX874913 Sequence
43	618	29.1	110000	2	AP006501	AP006501 Contination (9) of
44	577.5	27.2	26021	1	AE017258	AE017258 Wolbachia
45	568.5	26.8	333800	1	SM591792	SM591792 Stenothria

ALIGNMENTS

RESULT 1	BC011904	1908 bp	mRNA	linear	PRI 23-DEC-2003
LOCUS	BC011904				
DEFINITION	Homo sapiens O-sialoglycoprotein endopeptidase-like 1, mRNA (CDNA				
ACCESSION	BC011904.2	GI:40225818			
VERSION	BC011904				
KEYWORDS	MGC.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
	1 (bases 1 to 1908)				
	Strausberg, R.L., Pelngold, B.A., Grouse, L.H., Derge, J.G.,				
	Klausner, R.D., Collins, P.S., Wagner, C.M., Shennan, C.F., Bhat, N.K.,				
	Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,				
	Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Heib, F.,				
	Diachenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,				
	Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,				
	Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,				
	Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,				
	Abramson, R.D., Mullany, S.J., Bosak, S.A., McKwan, P.J.,				
	McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,				
	Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hilly, S.W.,				
	Villalón, D.K., Muzny, D.M., Sodergren, B.J., Lu, X., Gibbs, R.A.,				
	Fahy, J., Helton, B., Ketterman, M., Madan, A., Rodriguez, S.,				
	Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,				
	Boutfard, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D.,				

QY 321 LeuValAlaSerGlyValAlaSerAspPheTyrIleArgArgAlaLeuGluIleLeu 340
Db 1064 CTGGTTCGACATCGTGAGTGTCCGAAAGTAACTTATATCCGAGACCTCGGAAATTTTA 1123
QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeuCysThrAspAsnGly 360
Db 1124 ACAAGCGACACAGCGACCTTGTGTCTCTCCCGAGACTAGCACTGATATATGGC 1183
QY 361 IleMetIleAlaTTPAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380
Db 1184 ATTATGATTCGACGAAATGATGATGAAAGACTACGCTGCTGGGCACTTTTACATGAC 1243
QY 381 IleGluGlyIleArgTyrGluProLeuCysProLeuGlyValAspIleSerLeuGluVal 400
Db 1244 ATAGAGGCATCCGCTATTAACCAAAATGTCTTGGAGTAGACATATCAAAAGAAATT 1303
QY 401 GlyGluAlaSerIleLeuValProGluLeuLeuMetGluIle 414
Db 1304 GGAGAAAGCTTCCTAAAGTACCAATTAATAATGAGATA 1345
RESULT 2
AR428803 2197 bp DNA linear PAT 18-DEC-2003
LOCUS AR428803 Sequence 1 from patent US 6642041.
ACCESSION AR428803
VERSION AR428803.1 GI:40188589
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2197)
AUTHORS Chen,J., Feder,J.N., Nelson,T.C., Krystek,S.R. and Duclos,F.
TITLE Polynucleotides encoding a novel metalloprotease, MP-1
JOURNAL Patent: US 6642041-A 1 04-NOV-2003;
FEATURES
LOCATION/Qualifiers
SOURCE 1..2197
/organism="Unknown"
/mol_type="genomic DNA"
ORIGIN
Alignment Scores:
Pred. No.: 7,37e-171 Length: 2197
Score: 2125.00 Matches: 414
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0
DB: 6
US-10-649-273-2 (1-414) x AR428803 (1-2197)
QY 1 MetLeuIleLeuThrIleThrAlaGlyValPhePheLeuProSerLeuArgLeuValTyr 20
Db 231 ATGCTAACTTCTAGCTAAGACTGACGAGATTTTAAACCAACCAAAAGAAAGTTTAT 290
QY 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisIleValLeu 40
Db 291 GAATTTTAAAGAGTTTAAATTTTCACTCTGGAACACTATTTCTTCAATAAAATAGATTTG 350
QY 41 GlyIleGluThrSerCysAspAspThrAlaAlaAlaValAspGluThrGlyAsnVal 60
Db 351 GGAATTTGAAACTAGTGTGATGATACAGCAGCTGCTGTGGATGAACCTGGAATGTGC 410
QY 61 LeuGlyGluAlaIleHisSerGlnThrGluValHisLeuLeuThrGlyIleValPro 80
Db 411 TTGGAGAGAGCAATATCATTTCCAAACTGAAGTTCAATTTAAAAACAGGTGGATTTGCTT 470
QY 81 ProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100
Db 471 CCAGAGCTCAACAGCTTCAACAGAAATATTTCAACGATATGACAAAGAGCTCTTTCT 530
QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLeuProGlyLeu 120
Db 531 GCCAGTGGAGTCTCTCCAAAGTACCTCTCAGCAATTGCAACTAATAAACAGAGACTT 590

QY 121 AlaLeuSerLeuGlyValAlaGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLeu 140
Db 591 GCTTTAAGCCTGGAGAGGGCTTATCATTTAGCTTACAGCTGGTAGAGCAAGTTAAAAAG 650
QY 141 ProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLeu 160
Db 651 CCAATTCATCCCAATTCATATGAGAGCTCATGCACTTACTATTAGGTGACCAATTA 710
QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlnHisCysLeuLeuAlaLeuVal 180
Db 711 GTAGAAATTCCTTTTATAGTCTTTTATTTGATTTCTGGAAGTCACTGTCTGTGGCAATTGTT 770
QY 181 GlnGlyValSerAspPheLeuLeuLeuGlyValSerLeuAspIleAlaProGlyAspMet 200
Db 771 CAAGAGATTTCAGATTTCCTTCTTGGTAAAGCTTTGGACATATGACACGAGTGACATG 830
QY 201 LeuAspLeuValAlaArgArgLeuSerLeuIleHisIleProGluCysSerThrMetSer 220
Db 831 CTGACAAAGGTGGCAAGAACCTTTCTTAAATAAACAATCCAGAGTCTCCACCATAGAT 890
QY 221 GlyGlyValAlaIleGluHisLeuAlaValGlnGlyAsnArgPheHisAspAspIleLeu 240
Db 891 GGTGGGAAAGCCATAGAACATTTGGCCAAACAGAAATAGATTTTATTTGACATCAA 950
QY 241 ProProLeuHisHisIleAlaValAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260
Db 951 CTCCTTCGATCATGCTAAATAATTTGATTTTCTTTACTGGACCTTCAACAGCTTACT 1010
QY 261 AspLeuIleIleIleMetLeuValGluValGluGluGlyIleGluValGlnIleLeuSer 280
Db 1011 GATTAATAATAATAATGAAAAAGAAAAAGAGAAAGTTTGAAGAGGGCAAAATCTCTCT 1070
QY 281 SerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCysHisLeuValIleArg 300
Db 1071 TCAGCAGACAGACATTTGCTGCCAGATGACAGACCAATGGCATCTCTTTGTAAAGA 1130
QY 301 ThrHisArgAlaIleLeuPheCysValArgAspLeuLeuProGlnAsnAsnAlaVal 320
Db 1131 ACACATCGGGCTATTTCTGTTTGTAAAGCAGAGACTTTTACTCCAAATAATAGCAGTA 1190
QY 321 LeuValAlaSerGlyValAlaSerAspPheTyrIleArgArgAlaLeuGluIleLeu 340
Db 1191 CTGGTTCATCTGTGTGTGTGCGCAAGTAATCTTATATCCGAGACCTCGGAAATTTTA 1250
QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeuCysThrAspAsnGly 360
Db 1251 ACAAGCGACACAGCTGACCTTGTGTGTCTCTCTCCAGACTATGACACTGATTAATGGC 1310
QY 361 IleMetIleAlaTTPAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380
Db 1311 ATTATGATTCGATCGAATGATATTTGAAGACTAGCTGCTGGGCACTTTTACATGAC 1370
QY 381 IleGluGlyIleArgTyrGluProLeuCysProLeuGlyValAspIleSerLeuGluVal 400
Db 1371 ATAGAGGCATCCGCTATGAAACCAAAATGTCTCTTGAAGTAGACATATCAAAAGAAATT 1430
QY 401 GlyGluAlaSerIleLeuValProGluLeuLeuMetGluIle 414
Db 1431 GGAGAAAGCTTCATTAATAAGTACCAATTAATAATGAGATA 1472
RESULT 3
AR428808 1387 bp DNA linear PAT 18-DEC-2003
LOCUS AR428808 Sequence 21 from patent US 6642041.
ACCESSION AR428808
VERSION AR428808.1 GI:40188594
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1387)
AUTHORS Chen,J., Feder,J.N., Nelson,T.C., Krystek,S.R. and Duclos,F.

TITLE Polynucleotides encoding a novel metalloprotease, MP-1
JOURNAL Patent: US 662041-A 21 04-NOV-2003;
FEATURES Location/Qualifiers
SOURCE 1..1387
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN

Alignment Scores:

Score: 3.54e-168 Length: 1387
Score: 2090.50 Matches: 412
Percent Similarity: 93.85% Conservative: 0
Best Local Similarity: 93.85% Mismatches: 2
Query Match: 98.38% Indels: 25
DB: Gaps: 1

US-10-649-273-2 (1-414) x AK428808 (1-1387)

Qy 1 MetleuileleuthrlysthrAlaGlyValPhePheLysProSerLysArgLysValTyr 20
Db 24 ATGCTAACTTACATACAGCTGAGAGCTTTTAAACCATCAAGAAAGAAAGTTAT 83
Qy 21 GlupheLysArgSerPheAsnPhelSProGlyThrLeuPheLysLysValLeu 40
Db 84 GAATTTTAAGAAGTTTATTTTCTCTGAAACACATATTTCTCATATAAATAGTATTG 143
Qy 41 GYIIleGluThrSerCysAspAspThrAlaAlaValValAspGluThrGlyAsnVal 60
Db 144 GGAATTTGAAGCTGATGATGATACAGACGCTGCTGGTGGATGAAACTGGAATATGTG 203
Qy 61 LeuGlyValAlaIleHisSerGlnThrGlyValHisLeuLysThrGlyGlyIleValPro 80
Db 204 TTGGAGAAAGCAATATATTCCTCCAACTGAACTTATTAACACAGGTGGAGTTGTTCT 263
Qy 81 ProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100
Db 264 CCAGCAGCTCAACAGCTTCAACAGAGAAATATTCAGAGATATATCAAGAAAGCTTTCT 323
Qy 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeu 120
Db 324 GCCAGTGGAGTCTCTCCAACTGACCTCTCAGCAATTGCAACTCAATMAAACAGGACTT 383
Qy 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLysLys 140
Db 384 GCTTTAAGCTTGAGAGTGGGCTTATCATTTAGCTTACAGCTGGTAGCAGCTTAAATAAG 443
Qy 141 ProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLys 160
Db 444 CCATTCATTCCTCATTCATATAGAGGCTCATGCACTTATCTATTAGTTGACCAATMAA 503
Qy 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyYHIScyLeuLeuAlaLeuVal 180
Db 504 GTAGAAATTTCCCTTTTATGTTCTTTGATTTCTGAGAGTCACTGCTGTGGCATTTAGTT 563
Qy 181 GlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet 200
Db 564 CAAGGAGTTTCAATATTTCTGCTTCTTGAAGCTTTTGAACATAGCAACAGGAGCATG 623
Qy 201 LeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGluCysSerThrMetSer 220
Db 624 CTTTGACAAAGTGGCAAGAACCTTTCTTTAATAAACAATCCAGAGTCTCCACCATAGT 683
Qy 221 GlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLys 240
Db 684 GGGGGAAGAAAGCATTAAGCATTTGGCCAAACAGAAATAGATTTCATTTTGAACATCAA 743
Qy 241 ProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260
Db 744 CCTCCCTTGATCATGCTAATAAATTTGATTTTCTTTTACTGACCTTCAACAGTTACT 803
Qy 261 AspLysIleIleMetLysLysGlyLysGluGlnGlyIle----- 273
Db 804 GATTAATAATATATGAAAAAGAAAAAGAGAGATATATTTCTAATTAAGTAAAGTTGAA 863

Qy 274 -----GluLys 275
Db 864 CAGATTAATATCTCGGATTTGCTCTTAAATAAGCTGCTCATTTCTCAGGTATAGAGAG 923
Qy 276 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 295
Db 924 GGGCAAAATCTGCTTTCAGACAGACATGCTGCCACAGTACAGCACACATAGGCATGT 983
Qy 296 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnAspLeuLeuPro 315
Db 984 CATCTGTGAAGAAAGACACATCGGGCTATTTCTGTTTGTAAAGCAGAGACTTTTACT 1043
Qy 316 GlnAspAsnAlaValLeuValAlaSerGlyGlyValAlaSerAspPheTyrIleArgArg 335
Db 1044 CAAATATATCCAGTACGTGTGATCTGGTGTGTCTCCAGTACTTCTATATCCGACGA 1103
Qy 336 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeu 355
Db 1104 GCTCTGGAATTTTAAACAAACGCAACACAGTGCCTTGTGTCTCTCCCTCCAGACTA 1163
Qy 356 CysThrAspAsnGlyIleMetIleAlaThrAsnGlyIleGluArgLeuArgAlaGlyLeu 375
Db 1164 TGACATGATATATGCTATTAATGATTCATGAGATGATTTGAAGACTACGTGTGCTTG 1223
Qy 376 GlyIleLeuHisAspIleGlnGlyIleArgTyrGluProLysCysProLeuGlyValAsp 395
Db 1224 GGCATTTTACATGACATAGAAAGCATCGCTATGACCAAAATGCTCTTGAGATAGC 1283
Qy 396 IleSerLysGluValGlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 414
Db 1284 ATATCAAAAGAGTTGAGAAAGCTTCATATAAAGTACACAAATTAATAATGAGATA 1340

RESULT 4
HSA295148 1387 bp mRNA linear PRI 30-OCT-2000
LOCUS Homo sapiens mRNA for putative sialoglycoprotease type 2.
DEFINITION AJ295148
ACCESSION AJ295148.1 GI:11071726
KEYWORDS metalloproteinase; sialoglycoprotease.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1
Chen,J.M., Fortunato,M. and Barrett,A.J.
Cloning and sequencing of a second human putative
sialoglycoprotease homologue
Unpublished
2 (bases 1 to 1387)
AUTHORS Chen,J.M.
TITLE Direct Submission
JOURNAL Submitted (27-OCT-2000) Chen J.M., MRC Molecular Enzymology
Laboratory, The Babraham Institute, Babraham, Cambridge, CB2 4AT,
UNITED KINGDOM

FEATURES
source location/Qualifiers
1..1387
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue type="uterus"
24..1343
/function="metalloproteinase of family M22"
/codon_start=1
/product="putative sialoglycoprotease type 2"
/protein_id="CAC14666.1"
/db_xref="GI:11071727"
/db_xref="GOA:O9H4B0"
/db_xref="UniProt/TREMBL:O9H4B0"
/translation="MLITKTAGVFFKSKKRVFFLRSFNFPETTLPLAKIVLGIET
SCDDTAADVDETNVLEGAHSQTEVHLKTGGIVPPAAQQLHRENIQRIQVEALSAS
GVSPDSIAITTKIPGLAISVGLSFLDVGQLKRPFIPIHMEAHALTRITLNK
VEFPFVLIIISGHCILLALVQGVSDPFLLLGSLDIPADMIDKVARRLSLIKPFCST

MSGKAIEHLAKOGRPFHDIKPEPLHNAKCDSPFNGLOHTDKLIMKKEKRGFLI
SKVEQINIPGLCIKIAHFRERKQIILSSADIAITVOTPHACHVETHTRAILFECK
QRDLPPONNAVLASGVASNTYIRRAHILTNATQCTILCPPLRCTDNGIMIAWNG
IERLRGIGIILHIDIEGIRBPCKPLGVDISKEVGASIVPOLKMEI"

ORIGIN

Alignment Scores:

Pred. No.:	3,54e-168	Length:	1387
Score:	2090.50	Matches:	412
Percent Similarity:	93.85%	Conservative:	0
Best Local Similarity:	93.85%	Mismatches:	2
Query Match:	98.38%	Indels:	25
DB:	9	Gaps:	1

US-10-649-273-2 (1-414) x HSA295148 (1-1387)

```

QY 1 MetLeuIleLeuThyThrAlaGlyValPhePheLysProSerLysArgLysValTyr 20
DB 24 ATGCTAATCTTGACCTAAGACTGACAGAGTTTCTTTTAAACCATCAAAAAGAAAGTTTAT 83
QY 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisLysIleValLeu 40
DB 84 GAATTTTAAAGAGTTTAAATTTTTCATCTGAAACCTAATTTCTTCATTAATAAGTATTTG 143
QY 41 GlyIleGluThrSerCysAspAspThrAlaAlaValValAspGluThrGlyAsnVal 60
DB 144 GGAATTTGAACCTAGTTGTGATATACAGACCTGCTGTGTGATGAACTGGAAATGTG 203
QY 61 LeuGlyGluAlaIleHisSerGlnThrGlyValHisLeuLeuThyGlyIleValPro 80
DB 204 TTGGAGAAAGCAATACATTTCCAAACTGAAAGTTCATTTAAACAGGTGGAGTTGTCCT 263
QY 81 ProAlaAlaGlnGlnLeuHisArgGlyLysAsnIleGlnArgIleValGlnGlnLysLeu 100
DB 264 CCACAGAGCTCAACAGCTTCACAGAGAAATATTCACAGAAATGACAAAGAGCTCTTTCT 323
QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeu 120
DB 324 GCCAGTGGAGTCTCTCCAAAGTGAACCTCTCAGCAATGCACTAATCAACCAAGAGACTT 383
QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLysLys 140
DB 384 GCTTTAAACCTGGAGGTGGCTTATCATTTTACCTTACACGTGTGGAGAGTTAAAAAG 443
QY 141 ProPheIleProIleHisIleMetGlyAlaHisAlaLeuThrIleArgLeuThrAsnLys 160
DB 444 CCATTCATTTCCCATTCATCATATGAGAGGCTCATGCACTTACTATTTAGTTGACCAATTA 503
QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180
DB 504 GTAGAAATTTCTTTTATTTAGTTCTTTTGAATTCCTGAGAGGTCACTGTCTGTGGCATTAGTT 563
QY 181 GlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet 200
DB 564 CAAGAGATTTCAATTTCTGCTCTTGAAAGTCTTTGCAATGCAACAGAGTGCATG 623
QY 201 LeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlnCysSerThrMetSer 220
DB 624 CTTTACAAAGTGGCAAGAACTTTCTTTAATTAACAATCCAGAGGCTCCACCAAGTACT 683
QY 221 GlyGlyLysAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLys 240
DB 684 GTGGGAAAGCCATAGACATTTTGGCCAAACAAAGAAATAGATTTTCATTTTACATCAAA 743
QY 241 ProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260
DB 744 CCTCCCTTGACATCATGCTAAAAATGTGATTTTCTTTTACGCACTTCAACACGTTACT 803
QY 261 AspLysIleIleMetLysLysGlyLysGluGluGlyIle----- 273
DB 804 GATTAATAATTAATGAAGAAAGAAAGAGAGGATATTTCTTAATTAGTAAGTTGA 863
QY 274 -----GluLys 275

```

```

DB 864 CAGATTAATATTTCTGCGATTGTGCTTAAAAATACCTGCTATTCTGCAAGTATGAGAG 923
QY 276 GlyGlnIleLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCys 295
DB 924 GGGCAATCTGCTTTCAGCAGAGACCATTTGCTGCCACATGACACACAAATGGCATGT 983
QY 296 HisLeuValLysArgThrHisAlaGalaIleLeuPheCysLysGlnArgAspLeuLeuPro 315
DB 984 CATCTGTGAAAGAAACACATCGGCTTATTTCTGTTTGTAAAGCAGAGACTGTACTT 1043
QY 316 GluAspAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheThrIleArgArg 335
DB 1044 CAAAATTAATCAGACTGTTGTGATCTGTGTGTGTCCAAAGTAACTTTATATCCGAGA 1103
QY 336 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 355
DB 1104 GCTCTGGAATTTTAAACAAACGCAACAGATGCACTTTGTTGTCTCTCTCCAGACTA 1163
QY 356 CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeu 375
DB 1164 TGCACGTATATGCGCATTAATGATGATGGAATGATGTAAGACCTAATCGTGGGCTTG 1223
QY 376 GlyIleLeuHisAspIleGluGlyIleArgTyrGluProLysCysProLeuGlyValAsp 395
DB 1224 GGCATTTTACATGACATAGAAAGCATCCGCTATGAAACCAAAATGCTCTTGGAGTAGAC 1283
QY 396 IleSerIleGlyValAlaGlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 414
DB 1284 ATATCAAAAGAGTTGGAGAGCTTCCATTAAGTACCAACATTTAAAAATGAGATA 1340

RESULT 5
AX664697 1245 bp DNA linear PAT 22-MAR-2003
LOCUS AX664697
DEFINITION Sequence 6 from Patent WO2074960.
ACCESSION AX664697
VERSION AX664697.1 GI:29164457
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Leiby,K.R., Kapeller-libermann,R. and Glucksmann,M.
TITLE 38650, 28472, 5495, 65507, 81588 and 14354 methods and compositions
of human proteins and uses thereof
JOURNAL Patent: WO 02074960-A, 6 26-SEP-2002;
Millennium Pharmaceuticals, Inc. (US)
FEATURES
source 1..1245
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Alignment Scores:
Pred. No.: 5.03e-168 Length: 1245
Score: 2088.00 Matches: 407
Percent Similarity: 99.03% Conservative: 3
Best Local Similarity: 98.31% Mismatches: 4
Query Match: 98.26% Indels: 0
DB: 6 Gaps: 0

US-10-649-273-2 (1-414) x AX664697 (1-1245)
QY 1 MetLeuIleLeuThyThrAlaGlyValPhePheLysProSerLysArgLysValTyr 20
DB 1 ATGCTAATCTTGACCTAAGACTGACAGAGTTTCTTTTAAACCATCAAAAAGAAAGTTTAT 60
QY 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisLysIleValLeu 40
DB 61 GAATTTTAAAGAGTTTAAATTTTTCATCTGAAACCTAATTTCTTCATTAATAAGTATTTG 120

```

```

Qy 41 GlyIleGluThrSerCysAspAspThrAlaAlaValAspGluThrGlyAsnVal 60
Db 121 GGAATGAAACATGATTGATGATACAGAGCTGCTGTGTGATGAACTGGAATGTG 180
Qy 61 LeuGlyGluAlaIleHisSerGlnThrGluValHisLeuGlyThrGlyGlyIleValPro 80
Db 181 TTGGAGAGAAAGCATTCATTTCCCAACCTGAAGTTCATTTAAAAACAGGTGGATGTCTCT 240
Qy 81 ProAlaIleGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGlnAlaLeuSer 100
Db 241 CCAGCAGCTCAACACCTTCACAGAGAAATATTTCACGAAATGATCAAGAAAGCTCTTCT 300
Qy 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLeuProGlyLeu 120
Db 301 GCCAGTGAAGTCTCTCCAAAGTACCTCTCAGCAATTCGAACTAACATAAACAGAGACTT 360
Qy 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuVal 140
Db 361 GCTTAAGCCTGGAGTGGCTTATCATTTAGCTTACACCTGGTAGAGAGATTAAAAAG 420
Qy 141 ProPheIleProIleHisHisMetGluAlaHisAlaLeuThrIleArgLeuThrAsnVal 160
Db 421 CCATTCAATTCCTCATTCATCATATGAGAGCTCATGCACTTACTATTAAGTTGACCAATTA 480
Qy 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180
Db 481 GTAGAAATTCCTTTTATTGATCTTTTGAATTCCTGAGAGTCACTGTCTGTGGATTAAGT 540
Qy 181 GlnGlyValSerAspPheLeuLeuLeuGlyLeuSerLeuAspIleAlaProGlyAspMet 200
Db 541 CAAGAGATTTCAGATTTTCTGCTTCTGGAAAGTCTTTGACATACACAGAGTGCATG 600
Qy 201 LeuAspValValAlaArgArgLeuSerLeuIleValHisProGluCysSerThrMetSer 220
Db 601 CTGCAAGAGTGGCAAGAGACTTCTTAAATAAACATCCAGAGTCTCCACCATGAGT 660
Qy 221 GlyGlyValAlaIleGlnHisLeuAlaValGlnGlyAsnArgPheHisPheAspIleVal 240
Db 661 GGTGGAAAGCCATGAAACATTTGGCCCAACAGAAATGATTTGATTTGACATCAAA 720
Qy 241 ProProLeuHisHisAlaValAsnGlyAspPheSerPheThrGlyLeuGlnHisValThr 260
Db 721 CCTCCTTCATCATCATGCTAAATATGATGATTTTCTTTTACTGAGACTTCAACAGTTACT 780
Qy 261 AspValIleIleMetCysLeuGlyGluGlyGlyIleGlnValGlyGlnIleLeuSer 280
Db 781 GATTAATAATATGAAACAGAGAAACAGAGAGATGAGAAAGGGGCAATTCCTGTCT 840
Qy 281 SerAlaIleAspIleAlaIleAlaThrValGlnHisSerMetAlaCysHisLeuValVal 300
Db 841 TCAGCAGCAGACATTCCTCCACAGTACAGACACACATGCGATGCTCTTGAGAAAAGA 900
Qy 301 ThrHisArgAlaIleLeuPheCysLeuGlnArgAspLeuLeuProGlnAsnAsnAlaVal 320
Db 901 ACAACATCGGGCTATTCGTGTTTGTAGCAGAGAGACTGTTTACTCTCAAAATTAATGCA 960
Qy 321 LeuValAlaSerGlyValAlaAsnAsnPheThrIleArgArgAlaLeuGluIleLeu 340
Db 961 CTGGTTGACATCTGGTGTGTGCGCAAGTAACTTCTATATCCGACAGCTCGGAAATTTTA 1020
Qy 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeuCysThrAsnAsnGly 360
Db 1021 ACAAAACGCAACAGAGCACTTGTGTGTCTCTCCACAGACTATGCACTGATTAATGSC 1080
Qy 361 IleMetIleAlaThrAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380
Db 1081 ATTTAATGATTGCATGAAATGATTAAGAAACATACGCTGCTGGCTTTCATTAATGATAC 1140
Qy 381 IleGluGlyIleArgGlyGluProGlyCysProLeuGlyValAspIleSerValGluVal 400
Db 1141 ATAGAAGGATCCGCTATGAAACAAATATGCTCTTGAGATGACATATCAAAAGAGTT 1200
Qy 401 GlyGluAlaSerIleValProGlnLeuValMetGluIle 414

```

```

Db 1201 GGAGAGCTTCCTAAAGTACCAATTAATAATGAGATA 1242

RESULT 6
AX664695 1820 bp DNA linear PAT 22-MAR-2003
LOCUS AX664695
DEFINITION Sequence 4 from Patent WO2074960.
ACCESSION AX664695
VERSION AX664695.1 GI:29164455
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
1 Leiby, K.R., Kapeller-Libermann, R. and Glucksmann, M.
38650, 28472, 5495, 65507, 81588 and 14354 methods and compositions
of human proteins and uses thereof
Patent: WO 02074960-A 4 26-SEP-2002;
Millennium Pharmaceuticals, Inc. (US)

JOURNAL Location/Qualifiers
1. 1820
FEATURES
source
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
146..1390
/note="unnamed protein product"
/codon_start=1
/protein_id="CAD80044.1"
/db_xref="GI:29164456"
/translation="MDILKTAVVFEPKSRKRVYBFLRSFNHPTLFLHKVLGIET
SCDDTAAVVDERTGVLEAIHSQTEVHLRTGIVPAPAOIHRENIORIVQEAISAS
GVSPDLISATTIKPGALSLGVGSFSLQVGOLKKPFIPIHMEAHALITRLTK
VERPRVILISGHCLALVQVSDPLTGLKSLDIAFGMDLKVARRLSLKHPEGST
MSGKAIIEHLAKGNRPHDIKPRILHAKRCDSPFGLQVTPNNRNROEGIEG
QILSSADIALAIVQHTMACHLVTRTRAILFCQRDLIPQNNVAVASGGVASFYIR
RALEIITNAOCTLLCPRLCTDNGIMLAWNGIERIRAGILHDIIEGLRYEPCPL
GVDSKEVGBASIKVPOLKMEI"

ORIGIN
Alignment Scores:
Pred. No.: 8,14e-168 Length: 1820
Score: 2088.00 Matches: 407
Percent Similarity: 99.03% Conservative: 3
Best Local Similarity: 98.31% Mismatches: 4
Query Match: 98.26% Indels: 0
Gaps: 0
DB: 6

US-10-649-273-2 (1-414) x AX664695 (1-1820)

Qy 1 MetLeuIleLeuThrHisThrAlaGlyValPhePheLeuProSerLeuValValTyr 20
Db 146 ATGCTAACTTGAATAAGCTGCGAGGAGTTTATTAACCATCAAAAGGAAAGTTTAT 205
Qy 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisValIleValLeu 40
Db 206 GAATTTTAAAGAGTTTAAATTTTTCATCTCGGAACATATTTCTTCATAAATATGATGTG 265
Qy 206 GAATTTTAAAGAGTTTAAATTTTTCATCTCGGAACATATTTCTTCATAAATATGATGTG 265
Qy 41 GlyIleGluThrSerCysAspAspThrAlaAlaValAspGluThrGlyAsnVal 60
Db 266 GGAATTAACATAGTGTGATGATGATACAGAGCTCTGTGTGTGATGAAAGTGAATGTG 325
Qy 61 LeuGlyGluAlaIleHisSerGlnThrGluValHisLeuValThrGlyGlyIleValPro 80
Db 326 TTGGAGAGCAACAAATTCCTCAAACTGAAGTTCAATTAATAACAGGTGGAGATTGTCT 385
Qy 81 ProAlaIleGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGlnAlaLeuSer 100
Db 386 CCAGCAGCTCAACAGCTTCACAGAGAAATATTTCACAGAAATGATCAAGAAAGCTTCTCT 445
Qy 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLeuProGlyLeu 120
Db 446 GCCAGTGAAGTCTCTCCAAAGTACCTCTCAGCAATTCGAACTATCACTAAACAGAGACTT 505

```


QY 121 AlaleuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuValGly 140
 Db 506 GCTTAAAGCTGGAGTGGCTTATCATTTAGCTTACAGCTGGTAGACAGTTAAAAAG 565
 QY 141 ProPheIleProIleHisMetGluAlaHisAlaLeuThrIleArgLeuThrAspLeu 160
 Db 566 CCATTCACTCCATTCACTATGATGAGGCTCATGCACTTACTATAGGTGACCAATAAA 625
 QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180
 Db 626 GTAGAAATTTCTTTTGTAGTTCTTTGATTTCGAGAGTCACGTCTGTGGCATTAAGT 665
 QY 181 GlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet 200
 Db 686 CAAGAGTTTCAGATTTCCTGCTTCTTGGAAGTCTTTGGACATAGACACAGGTGACATG 745
 QY 201 LeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGluCysSerThrMetSer 220
 Db 746 CTTGACAAAGTGGCAAGAAAGACTTTCTTAATATAAACATCCAGAGTCTCCACCATGAGT 805
 QY 221 GlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLys 240
 Db 806 GGTGGGAAAGCCATAGAACATTTGGCCAAACAAAGAAATGATTTCATTTTGACATCAAA 865
 QY 241 ProProLeuHisHisAlaLysAsnGlyAspPheSerPheThrGlyLeuGlnHisValThr 260
 Db 866 CTTCCCTTGATCATGCTTAAATATGTAATTTTCTTTTACTGACTGCAACAGTTACT 925
 QY 261 AspLysIleIleMetLysLysGlyLysGlyGlyGlyIleGlyLysGlyGlnIleLeuSer 280
 Db 926 GATAAAATTAATGAAAAACAGAAACAAAGAAATGAAAGGCGCAAAATCTGTCT 985
 QY 281 SerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCysHisLeuValLysArg 300
 Db 986 TCAGACAGCAGCATTTGCTCCACAGTACGACACACATGCGATGCTTGTGAAAGA 1045
 QY 301 ThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuProGlnAsnAlaVal 320
 Db 1046 ACACATCGGGCTATTCTGTGTTGTAGACAGAGAGCTTTGACTCAAAATATGACAGTA 1105
 QY 321 LeuValAlaSerGlyGlyValAlaAsnAsnPheTyrlleArgArgAlaLeuGlnIleLeu 340
 Db 1106 CTGTTGATCATCTGGTGTGCTCCAGTAACTTCTATATCCGACAGACTCGAAATTTTA 1165
 QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeuGlySerThrAspGly 360
 Db 1166 ACAACGCAACACAGTGCATTTGTGTCTCTCCCACTAGCACTGATTAATGAGC 1225
 QY 361 IleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380
 Db 1226 ATTAAGATTGCAATGAAATGATTAAGAAACATACGCTGGCTGGCATTTTACATGAC 1285
 QY 381 IleGluGlyIleArgTyrlleProLysCysProLeuGlyValAspIleSerLysGlyVal 400
 Db 1286 ATAGAGGCATCGCTATGAAACCAAAATGTCTCTTGAGTGAACATATCAAAAGAAATT 1345
 QY 401 GlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 414
 Db 1346 GAGAGAGCTTCATTAAGTACCACAATTAATAATGAGATA 1387
 RESULT 7
 AX713716 2208 bp DNA linear PAT 15-APR-2003
 LOCUS AX713716
 DEFINITION Sequence 400 from Patent EPI293569.
 ACCESSION AX713716
 VERSION AX713716.1 GI:29888642
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1

AUTHORS Isogai, T., Sugiyama, T., Otsuki, T., Wakamatsu, A., Sato, H., Ishii, S., Yamamoto, J. T., Isono, Y., Hio, Y., Otsuka, K., Nagai, K., Irie, R., Tamechika, I., Seki, N., Yoshikawa, T., Otsuka, M., Nagahari, K. and Masuno, Y.
 TITLE Full-length cDNAs
 JOURNAL Patent: EP 1293569-A 400 19-MAR-2003;
 Helix Research Institute (JP) ; Research Association for Biotechnology (JP)
 FEATURES
 source location/Qualifiers
 1..2208
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"
 ORIGIN
 Alignment Scores:
 Pred. No.: 1,836-155 Length: 2208
 Score: 1944.00 Matches: 386
 Percent Similarity: 93.24% Conservative: 0
 Best Local Similarity: 93.24% Mismatches: 4
 Query Match: 91.48% Indels: 24
 DB: Gaps: 1
 US-10-649-273-2 (1-414) x AX713716 (1-2208)
 QY 1 MetLeuIleLeuThrLysThrAlaGlyValPhePheLysProSerLysArgLysValTyr 20
 Db 344 ATGTATATCTTGACTAAGATGCGAGAGTTTATTAACCATCAAAAGAAAGTTTAT 403
 QY 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLysLysIleValLeu 40
 Db 404 GAATTTTAAAGATTAAATTTTATTTTCATCTGGAACATGATTTCTTCAATAATAGTATG 463
 QY 41 GlyTllegLysSerCysAspAspThrAlaAlaValAlaValAspGluThrGlyAsnVal 60
 Db 464 GGAATGAAACCTAGTTGTGATGATACAGCAGCTGCTGTGATGAAACCTGAAATGTCG 523
 QY 61 LeuGlyValAlaIleHisSerGlnThrGlyValHisLeuLysThrGlyGlyIleValPro 80
 Db 524 TTGGAGAACCAATATCACTTCCCAACTGAAAGTTTATTAACAAAGTGGAGTTTCTCT 583
 QY 81 ProAlaAlaGlnGlnLeuHisArgGlyAsnIleGlnArgIleValGlnGluAlaLeuSer 100
 Db 584 CCAAGAGCTCAACAGCTTCAAGAGAAATATTAACGAATAGTACAAAGAAAGCTCTTCT 643
 QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeu 120
 Db 644 GCCAGTGAAGTCTCTCCAAAGTCACTCTCAGCAATTCACATCAATAAACCAAGACTT 703
 QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLysLys 140
 Db 704 GCTTAAAGCTGGAGAGGGCTTATCATTTAGCTTACAGCTGGAGAGCAGTTAAAAAG 763
 QY 141 ProPheIleProIleHisMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLys 160
 Db 764 CCATTCACTCCATTCACTATGAGAGCTCATGCACTTACTATAGGTGACCAATAAA 823
 QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180
 Db 824 GTAGAAATTTCTTTTGTAGTTCTTTGATTTCGAGAGTCACTGTCTGTGGCATTAAGT 883
 QY 181 GlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet 200
 Db 884 CAAGAGTTTCAATATTTCTGCTTCTTGAAAGCTTTGACATAGACCAAGGTGACATG 943
 QY 201 LeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGluCysSerThrMetSer 220
 Db 944 CTTGACAAAGTGGCAAGAGACTTCTTTAATATAACATCCAGAGTCTCCACCATGAGT 1003
 QY 221 GlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLys 240
 Db 1004 GGTGGGAAAGCCATAGAACATTTGGCCAAACAAAGAAATGATTTCATTTTGACATCAAA 1063

Qy 241 ProProLeuHisHSHaLaLyAsnCyAspPheSerPheThrGlyLeuGlnHisValThr 260
 Db 1064 CTTCCCTTCATCATGCTAAATATGTATTTCTTTTACCTGAGCTTCAACGTTACT 1123
 Qy 261 AspLysIleIleMetLysLysGlnLysGlnGlnGlnGlnGlnGlnGlnGlnGln 280
 Db 1124 GATAAATTAATTAATGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 1183
 Qy 281 SerAlaAlaAspIleAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 300
 Db 1184 TCACAGACAGACATTCCTGCGCACATGACACACATGACATGACATGACATGACAT 1243
 Qy 301 ThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuProGlnAsnAlaVal 320
 Db 1244 ACAATCGGGCTATTCCTGTTTGAAGCAGAGACTTTCCTCAAAATTAATGACAGA 1303
 Qy 321 LeuValAlaSerGlyGlyValAlaAlaSerAsnPheTyrIleArgArgAlaLeuGlnIle 340
 Db 1304 CTGGTTGCATCTGGTGTGTCGCAAGTAATCTTGATCCGACAGACTCTGGAATTTTA 1363
 Qy 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeuCysThrAspAsnGly 360
 Db 1364 ACAACGCAACACGACGACCTTTGTTGTCCTCTCCGACGACGACGACGACGACGAC 1423
 Qy 361 IleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeuGlnIleLeuHisAsp 380
 Db 1424 ATTAAGATTCG----- 1435
 Qy 381 IleGluGlyIleArgTyrGlnProLysCysProLeuGlyValAspIleSerLysGlnVal 400
 Db 1436 -----TGATGTCTCTTGAGAGTACATCAAAAGAAAGTT 1471
 Qy 401 GlyGluAlaSerTyrLysValProGlnLeuLysMetGlnIle 414
 Db 1472 GGAGAGAGCTTCATAAAGTACACATTAATAATGACAGATA 1513

RESULT 8
 AK055441
 LOCUS
 DEFINITION Homo sapiens cDNA FLJ30879 fls, clone FEBRA2004592, highly similar to Homo sapiens mRNA for putative sialoglycoproteinase type 2.
 AK055441
 ACCESSION
 VERSION AK055441.1 GI:16550166
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R., Wakamatsu, A., Hayashi, K., Sato, H., Nagai, K., Kimura, K., Makita, H., Sekine, M., Oobayashi, M., Nishi, T., Shibahara, T., Tanaka, T., Ieshi, S., Yamamoto, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y., Nagahara, K., Murakami, K., Yasuda, T., Iwayanagi, T., Wagatsuma, M., Shiratori, A., Sudo, H., Hosoiri, T., Kaku, Y., Kodaira, H., Kondo, H., Sugawara, M., Takahashi, M., Kanda, K., Tokoi, T., Furuya, T., Kikawa, E., Omura, Y., Abe, K., Kamihara, K., Katsuta, N., Sato, K., Tanikawa, M., Yamazaki, M., Ninomiya, K., Ishibashi, T., Yamashita, H., Murakawa, K., Fujimori, K., Tanai, H., Kimata, M., Watanabe, M., Hirooka, S., Chiba, Y., Iehida, S., Ono, Y., Takiguchi, S., Watanabe, S., Yosida, M., Horuta, T., Kusano, J., Kanehori, K., Takahashi-Fujii, A., Hara, H., Tanase, T., Nomura, Y., Togiya, S., Komai, F., Hara, R., Takeuchi, K., Arita, M., Imose, N., Musashino, K., Yuuki, H., Oshima, A., Sasaki, N., Aotsuka, S., Yoshikawa, Y., Matsunawa, H., Ichihara, T., Shihohara, Y., Sano, S., Moriya, S., Momiyama, H., Satoh, N., Takami, S., Terashima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H., Goto, Y., Shimizu, F., Wakebe, H., Hishigaki, H., Watanabe, T., Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K., Kumagai, A., Itakura, S., Fukumizu, Y., Fujimori, Y., Komiyama, M., Tashiro, H., Tanigami, A., Fujiwara, T., Ono, T., Yamada, K., Fujii, Y., Ozaki, K., Hiro, M., Ohmori, Y., Kawabata, A., Hikiji, T., Kobatake, N., Inagaki, H., Ikema, Y., Okamoto, S., Okitani, R., Kawakami, T., Noguchi, S., Itoh, T., Shigeta, K., Senba, T., Matsunura, K.,

REFERENCE
 JOURNAL
 TITLE
 AUTHORS
 Nishitani, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togashi, T., Oyama, M., Hata, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, J., Satoh, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K., Nagase, T., Nomura, N., Kikuchi, H., Masuh, Y., Yamashita, R., Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isogai, T., and Sugano, S. Complete sequencing and characterization of 21,243 full-length human cDNAs
 Nat. Genet. 36 (1), 40-45 (2004)
 2
 Nishi, T., Nakagawa, S., Senoh, A., Mizoguchi, H., Inagaki, H., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ieshi, S., Yamamoto, J., Isono, Y., Kawai, H., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsuno, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Magatsuma, M., Murakawa, K., Kanehori, K., Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahara, K., Masuh, Y., Nagai, K., and Isogai, T. NEDO human cDNA sequencing project
 Unpublished
 3 (bases 1 to 2208)
 Isogai, T., Otsuki, T., and Sugiyama, T.
 Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute, Genomics Laboratory, 1532-3 Yana, Kisarazu, Chiba 252-0812, Japan (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
 NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: RAB and HRI.

FEATURES

Source

1. 2208
 location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="FEBRA2004592"
 /cissue_type="brain"
 /clone_lib="FEBRA2"
 /dev_stage="fetus"
 /note="cloning vector: pMB18SFL3"
 344. 1438
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="BAB70923.1"
 /db_xref="GI:16550167"
 /translation="MLITKTAGVFPKSRKRYEFLRSPFHPGTLFLNKIVGIEITSCDDTAAYVDENGVIGEAISHQTEVHLKGGIVPPAAOOLHRENIORIVQALASGVSPDSLSAATITTKRGLASLGVGSFSLQVGLKKPPIPHHMHAAHTITLTKVERPFLVLLISGGICLALVQGSDFLLKSLDIADGMDLKYARLPLKHGECSTMSGGKILHAKGKRNRFDIKPLHAKKDCFFPTGLQHTVDKIIMKKEGEBTEGQIISASADIATVQHTWACHLVKTRTHRAILFCQKRDLLPQNNAVLVAAGVANSFCIRFALEILTNATQCTILCPPRLCTDNGIMIA"

CDS

ORIGIN

Alignment Scores:
 Pred. No.: 1,83e-155 Length: 2208
 Score: 1944.00 Matches: 386
 Percent Similarity: 93.24% Conservative: 0
 Best Local Similarity: 93.24% Mismatches: 4
 Query Match: 91.48% Indels: 24
 DB: 9 Gaps: 1

US-10-649-273-2 (1-414) x AK055441 (1-2208)

Qy 1 MetLeuIleLeuThrIleThrAlaGlyValAlaPhePheLysProSerIleArgLysValTyr 20
 Db 344 ATGCTAATCTTGACATGACGCGAGGATTTTAAACCATCAAAAGAAAGATTAT 403
 Qy 21 GluPheLysArgSerPheAsnPheIleProGlyThrLeuPheLeuHisIleValLeu 40

Db	404	GAATTTTAAAGAAGTTTAAATTTTCACTCGGAAACACTTTTCTTCAATAAATGATATG	463
Qy	41	GIYIIIEGIUThSeCYeAspAspRhrIaIaIaIaValAspGIUThrGIYAsnVal	60
Db	464	GGAAATTGAACTAAGTTGTGATGATATCAGCAGCGTGTGTGTGAATGAACATGGAAATGTG	523
Qy	61	LeuGIYGIUaIaIaIeHisSerGIInhrGIUaIhIeUeUeThrGIYGIYIIEVaIPro	80
Db	524	TTGGAGAAAGCAATATCTCCAAACTGAAGTTCACTTTAAAACAGTGGATTGTTCTT	583
Qy	81	ProIaIaIaGIInIeUhiSaRgGIuSnIIEGIaRgIIEVaIGInIuIaIeUSeR	100
Db	584	CCAGACAGCTCAACACACTTCACAGAGAAATATTCAACGAAATGATCAAGAGCTCTTCT	643
Qy	101	AlaSerGIYValSeRProSeRAspLeuSerAlaIeIaIaThrThIIElySeProGIYLeu	120
Db	644	GCCAGTAGAGTCTCTCCAAAGTACCTTCAGAAATGGCAACTACATAAACAGAGACTT	703
Qy	121	AlaIeUSeRLeuGIYValGIYLeuSeRPhSeRLeuGIInIeUaIaIGInIeUeUyS	140
Db	704	GCTTTAAAGCTGGAGTGGGCTTATTCATTACTTCAAGCTGTGTGAACAGTTAAAAAAG	763
Qy	141	ProPheIIEProIIEhISHeWecGIuAaIaIaIaIeUeThrIIEaRgLeuThRAsnY	160
Db	764	CCATTCACTCCATTCATCATATGAGGCTCATGCACTTACTATTAAGTTAGTTCACAAATA	823
Qy	161	ValGIUePheProPheLeuValIeUeUeIIESeRGIYhISCYeUeUeUaIeUaVal	180
Db	824	GTAGAAATTTCTTTTATTTTATTTCTTTGATTTCTGAGAGGTCACTGTCTGTGGCAATTAGT	883
Qy	181	GIInGIYValISeRAspPheIeUeUeUeUeGIYySeRLeuAspIIEaIaProGIYAspSe	200
Db	884	CAAGGAGTTCAAGATTTTCTGTCTTTGAAAGCTTTGGAAATGACACACAGTACATG	943
Qy	201	LeuAspYIIEIIEMeUeUeUeGIYySeRLeuIIElyhISProGIYCySeSeThRmetSe	220
Db	944	CTTGACAAAGTGGCAAGAAAGACTTCTTAAATAAACAATCAGAGTGTCTCCACATGACT	1000
Qy	221	GIYGIYySaIaIIEGIuHISeUaIaIeUaGIYySaMaRgPhehISpHeAspIIEly	240
Db	1004	GGTGGGAAGCCATGGAACATTTGGCCAAACAGGAATAATGATTCATTGTGACATCAAA	1063
Qy	241	ProPheUeUhiSHeIaIeUaSeNCyAspRPhSeSeRPhetThGIYLeuGIInISaIaTh	260
Db	1064	CTCTCCCTTGACATCAAGCTAAAAATTTGATTTTCTTTTATCGACATTCAACACGTTACT	1123
Qy	261	AspYIIEIIEMeUeUeUeGIYySeRGIuGIuGIYIIEGIUySaGIYGIInIIEUSeR	280
Db	1124	GATAAATATATATATAAAGAAAGAAAGAAAGAAAGATATTGACAAAGGGCAAAATCTGTCT	1183
Qy	281	SeRaIaIaAspIIEaIaIaIaThrValGIInhISThRmetIaCYhISeUeUaIySaRg	300
Db	1184	TCAGACAGACATGCTGCCACAGTACAGACACAAATGGCATGTCTGTGAAAAGA	1243
Qy	301	ThrHISaRgIaIaIIEUePheCYeUySGIuMaRgAspLeuProGIInAsnAaIaVal	320
Db	1244	ACACATCGCGGTATTCGTCTTTTGTAAAGACAGAGACTTTTAACTCAAAATATATCAGTA	1303
Qy	321	LeuValaIaSeRGIYValaIaSeRAsnPheryIIEaRgATaIaIeUeGIuIIEUe	340
Db	1304	CTGGTTCATCTGGTGTGTGCAAGTAACCTCTGTATCCGAGAGCTCTGTGAAATTTTA	1363
Qy	341	ThRaSnAaIaThRGIInCYSThrLeuUeUCySProProPolarGLeUCySThrAspAsnGI	360
Db	1364	ACAAACGCAACACAGTGCATTTGTGTGTCTCTCCACAGATATGCACTATATATGGC	1422
Qy	361	IIEweCIIeIaIaTPAsnGIYIIEGIuMaRgLeuMaRgIaGIYLeuGIYIIEUeUhiSa	380
Db	1424	ATTATGATTGCA-----	1433
Qy	381	IIIEGIuGIYIIEaRgTYGIUProLySeCYsProLeuGIYValaAspIIESeRlyeGIYVal	400
Db	1436	-----TGAATGCTCTCTTGGAGTAAGATATATCAAAAGAAAGCT	1471

[illegible]

source
1..1844
/organism="Mus musculus"
/mol_type="mRNA"
/strain="Czech 11"
/db_xref="taxon:10090"
/clone="MGC:67870 IMAGE:5012054"
/tissue_type="Mammary tumor metastasized to lung. Tumor arose spontaneously from a senescent normal mammary (clonal) outgrowth infected with the virus MMTV."
/clone_1ib="NCI_CGAP_Lu29"
/lab_host="DH108"
/note="Vector: pCMV-SPORT6"
222..1466
/codon_start=1
/product="Unknown (protein for MGC:67870)"
/protein_id="AAH58172.1"
/db_xref="GI:34849664"
/translation="MMLRRTAGAIKPKPKSKVYGLRFRSVPHTLSCHKLVGIET
SCDDGAIVDEGTAVGLSALHSOTVHLKTGIVPVVQDLRENIQRIVEETLSAS
RITBEDLSAIAATTIKGLALSLVGLSPSLQVWPKKPEFIHMEHALTRILNFK
VEPFLVLLISGGHCLALVQVSDPILGSLDIAKGMILKVARSLIATGPEST
MSGKAIHSLADGNRFHTINPQNAKNDPSTGLQHTIDKLITHEKSEIKG
QILSSADIAAVQATACHLAKRTHRALIFCKQNLSPANAVLVSGVASMLYR
KALRIVANATQTLCPPEPLCTDNGIMIANMGIERLRAGLGVLDHVEDIRBPKPL
GIDISREVAEAAIKVPRLMAL"
333..1397
/note="QRI7, Region: Metal-dependent proteases with possible chaperone activity (Posttranslational modification, protein turnover, chaperones)"
/db_xref="CDD:COG0533"
misc_feature
Alignment Scores:
Pred. No.: 2.71e-146 Length: 1844
Score: 1835.00 Matches: 352
Percent Similarity: 91.06% Conservative: 25
Best Local Similarity: 85.02% Mismatches: 37
Query Match: 86.35% Indels: 0
DB: 10 Gaps: 0
US-10-649-273-2 (1-414) x BC058172 (1-1844)
QY 1 MetLeuIIleuThrIlyThrIlaGlyValPhePheIysProSerIysArgIysValIyr 20
DB 222 ATGCTAAAGTTTAAAGAAACAGACGAGCTATTCCCAAGGCCCAAGAGTAAGTTAT 281
QY 21 GluPheLeuArgSerPheAsnPhenHisProGlyThrLeuPheLeuHisIlyLeuVal 40
DB 282 GGATTTTAAAGATTAGTGTTCATCCAGAACTCTCTTGTCTATAAAGTGTCG 341
QY 41 GlyIleGluThrSerCysAspPheThrIlaIaIaValValAspGluThrGlyAsnVal 60
DB 342 GGAATTGAAACAGCTGATGACACAGAGCGGCTGTGATGATAAATCGGAATGTG 401
QY 61 LeuGlyGluAlaIleHisSerGlnThrGluValHisIleLeuIysThrGlyIleValPro 80
DB 402 CTGGGGGAAGACATGCATCCCAACTCAGTTCACTGAAAACAGGTGGATGTTCTT 461
QY 81 ProAlaIaGlnGlnLeuHisAspGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100
DB 462 CCAAGTAGCTCAACACTTCCACAGAGAAATATTCAACGAATAGTAGAAGAACTTTCT 521
QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleIysProGlyLeu 120
DB 522 GCCAGTAGAATACCCCAAGCATCTCTGGCAATGCAACTACATCAAAACGGGACTG 581
QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuIysIys 140
DB 582 GCCCTTAAGCTTAGAGACTCGCTTATCTTTAGCTTACAGCTAGTAATTCGTTTAAAG 641
QY 141 PropheIleProIleHisIswetGluAlaHisAlaLeuThrIleArgLeuThrAsnIys 160
DB 642 CCGTTCATCCCGATCATCATGAGAGCTCAGCACTGACTATTTAGGCTCACCAATATA 701

QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180
DB 702 GTAGAATTTCTCTTTCTAGTCTTTTGAATTTCTGGGGGTACATGCTGTGGCATTAGTC 761
QY 181 GlnGlyValSerAspPheLeuLeuGlyIysSerLeuAspIleAlaProGlyAspMet 200
DB 762 CAAGGTGTTCCATTTCTCTGCTCTGGGAACTCTTTGACATAGGCGCAGCATG 821
QY 201 LeuAspIysValAlaArgArgLeuSerLeuIleIysHisProGluCysSerThrMetSer 220
DB 822 CTGACAGAGTGCAAGAACTTTCTTTATGACAAATCAGAAATTTTACAAATAGT 881
QY 221 GlyGlyValAlaIleGluHisLeuAlaIysGlnGlyAsnArgPheHisPheAspIleIys 240
DB 882 GCGGAAAGCTTATAGAACTTTGGCCAAAGAGAAATAGATTCATTTTACTATCAAT 941
QY 241 ProProLeuHisHisAlaIysAsnCyAspPheSerPheThrGlyLeuGlnHisValThr 260
DB 942 CCACCTAATGACAAATGCTAAGAAATTTGATTTTTCGGGACTTCAACATTAATCT 1001
QY 261 AspIysIleIleMetIysGlyIysGluGluGlyIleGluIysGlyGlnIleLeuSer 280
DB 1002 GATTAAGCTAATACACACAAAGAAAGAAAGCAATTAAGAAAGGGGCAATCTGTCA 1061
QY 281 SerAlaAlaAspIleAlaIleAlaThrValGlnHisThrMetAlaCysHisLeuValIysArg 300
DB 1062 TCAAGCTGACACATTTGCTGTGGGTACACATGCAACAGCTGCCACTTGCAGAAAAGA 1121
QY 301 ThrHisArgAlaIleLeuPheCysIysGlnArgAspLeuProGlnAsnAlaVal 320
DB 1122 ACACATCGTCTAATTTGTTTTCAGACAGAAATAATTTGCTAATCTCCAGCTAACGCACTA 1181
QY 321 LeuValAlaSerGlyGlyValAlaAspAsnPheIyrIleArgArgAlaLeuGluIleLeu 340
DB 1182 TTAGTTGTAATCTGAGAGTGTTCAGAACTTGTACATCCGAAAGCAATTGAAATGTCTC 1241
QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeuCysThrAspAsnGly 360
DB 1242 GCAAATGCAACGAGTGCACCTGCTGTGTCCTCCCTCCAGACTGTGCATGACAAATGCGC 1301
QY 361 IleMetIleAlaTrpAsnGlyIleGluArgLeuArgIaGlyLeuGlyIleLeuHisAsp 380
DB 1302 ATCATGATTCATGAGATGGAATGAAATTAAGTATGCTGTGGCTGTGGCTTTTACATGAT 1361
QY 381 IleGlnGlyIleArgIyGluProIyCysProLeuGlyValAspIleSerIysGluVal 400
DB 1362 GTAGAAGACATCCGATATGAACCAAAATGTCCTCTTGAAATAGACATATCCAGAAAGTT 1421
QY 401 GlyGluAlaSerIleIysValProGlnLeuIysMetGluIle 414
DB 1422 GCAGAAAGCTGCATTAAGTACCGGATTAATAATGCGACTT 1463
RESULT 10
AR541929 1416 bp DNA linear PAT 08-OCT-2004
LOCUS AR541929
DEFINITION Sequence 177 from patent US 6743619.
ACCESSION AR541929
VERSION AR541929.1 GI:53934009
KEYWORDS
SOURCE
ORGANISM
Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 1416)
Tang,Y.T., Zhou,P., Goodrich,R., Liu,C., Asundi,V., Ren,F.,
Zhang,J., Zhao,Q.A., Yang,Y., Xue,A.J., Wehrman,T., Wang,J.-R.,
Wang,D. and Dmanac,R.T.
Nucleic acids and polypeptides
Patent: US 6743619-A 177 01-JUN-2004;
FEATURES
source
1..1416
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN

Alignment Scores:

Pred. No.: 5.9e-139 Length: 1416
 Score: 1747.00 Matches: 340
 Percent Similarity: 100.00% Conservative: 1
 Best Local Similarity: 99.71% Mismatches: 0
 Query Match: 82.21% Indels: 0
 DB: 6 Gaps: 0

US-10-649-273-2 (1-414) x ARS41929 (1-1416)

QY 74 LysThrGlyGlyIleValProProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnAaG 93
 Db 280 AGAACAGGTGGAGATTGTTCTCCACAGCTCAACAGCTTCAACAGAAATATTTCAACCA 339
 QY 94 IleValGlnGlnAlaLeuSerAlaSerGlyValSerProSerAspSerAlaIleAla 113
 Db 340 ATAGTACAGAAAGCTCTTCTGCCAGTGGAGTCTCTCCAGTGAAGCTTCAGCAATTGCA 399
 QY 114 ThrThrIleLysProGlyLeuAlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGln 133
 Db 400 ACTACCAATAAACAGAGCTGCTTTAAGCTGGAGTGGCTTATCATTTAGCTTACAG 459
 QY 134 LeuValGlyGlnLeuLysLysProPheIleProIleHisIleMetGlnAlaHisAlaLeu 153
 Db 460 CTGGTAGGACAGATTAAAGAGCATTCATCCATTCATCATATGAGGCTCATGACATT 519
 QY 154 ThrIleArgLeuThrAsnLysValGluPheProPheLeuValLeuLeuIleSerGlyGly 173
 Db 520 ACTATAGGTGACCAATAAAGTAAATTTCTTTTAAATTTGATTTCTGAGAGT 579
 QY 174 HisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeu 193
 Db 580 CACTGCTGTTGGCATTTAGTTCACAGAGTTTCAGATTTTCTGCTTGGAAAGTCTTTG 639
 QY 194 AspIleAlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHis 213
 Db 640 GACATAGACACAGGAGCATGCTTGAACAAGTGGCAAGAAAGATTCTTTAATAAACAAT 699
 QY 214 ProGluCysSerThrMetSerGlyGlyLysAlaIleGlnHisLeuAlaLysGlnGlyAan 233
 Db 700 CCAAGAGTCTCCACCATAGTGGTGGAAAGCCATAGAACATTGTGCCCAACAGCAAAAT 759
 QY 234 ArgPheHisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPhe 253
 Db 760 AGATTTTCAATTTGACATCAAACTCTCTTGCATCAATGCTAAATAATTTGATTTCTTTT 819
 QY 254 ThrGlyLeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGlyIle 273
 Db 820 ACTGAGCTTCAACAGCTTACGTATTAATAATTAATGAAGAAAGAGAGAGATTT 879
 QY 274 GluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMet 293
 Db 880 GAGAGAGGGCAAACTCTCTTCAAGCACAGACATGTGCTGCAAGTACAGACACAAATG 939
 QY 294 AlaCysHisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeu 313
 Db 940 GCATGCTCATCTTGGAAAGAAACACATGGGCTATTTGTTTGAAGAGAGAGAGCTTG 999
 QY 314 LeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheTyrIle 333
 Db 1000 TTACTCTCAAAATTAATGCACTGCTGATTCATCTGATGCTGCAAGTAACTTCAATATC 1059
 QY 334 ArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLysCysProProPro 353
 Db 1060 CGCAGAGCTCTGGAATTTTAACAAACCAACACAGTGCACATTTGTGTGCTCTCTGCC 1119
 QY 354 ArgLeuGlySerThrAspAsnGlyIleMetIleAlaThrAsnGlyIleGluArgLeuArgAla 373
 Db 1120 AGACTATATCACTGATTAATGACATTAATGATTCATGAAATGATTAAGATACAGTCTCT 1179
 QY 374 GlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrGluProLysCysProLeuGly 393

Db 1180 GGCTGGGCAATTTTACATGACATAGAGGAGATCCGCTATGAGCAAAATGCTCTTGA 1239
 QY 394 ValAspIleSerLysGlyValGlyValAlaSerIleLysValProGlnLeuLysMetGlu 413
 Db 1240 GTAGACTATTCAAAGAGATTGAGAGAGCTTCCATTAAGTACCAATTAATAATGAG 1299
 QY 414 Tle 414
 Db 1300 ATA 1302
 RESULT 11
 BC078974
 LOCUS
 DEFINITION
 Rattus norvegicus cDNA clone IMAGE:1111906, partial cds.
 ACCESSION
 BC078974
 VERSION
 BC078974.1 GI:50926879
 KEYWORDS
 SOURCE
 ORGANISM
 Rattus norvegicus (Norway rat)
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 1546)
 Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
 Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
 Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
 Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
 Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
 Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
 Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
 Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
 Abramson,R.D., Mullany,S.J., Bosak,S.A., McEwan,P.J.,
 McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
 Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Huylk,S.W.,
 Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
 Fahey,J., Helton,E., Kettelman,M., Madan,A., Rodriguez,S.,
 Sanchez,A., Whitting,M., Madan,A., Young,A.C., Shevchenko,Y.,
 Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
 Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
 Butlerfield,Y.S., Krzywinski,M.I., Skalski,U., Smalls,D.B.,
 Schermer,A., Schein,J.E., Jones,S.J., and Marra,M.A.
 Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 12477932
 2 (bases 1 to 1546)
 DIRECTOR MGC Project.
 DIRECTOR MGC Project.
 DIRECTOR MGC Project.
 Submitted (02-AUG-2004) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Howard Jacobs
 CDNA Library Preparation: Express Genomics
 DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)
 Center, Stanford University School of Medicine, Stanford, CA 94305
 Web site: <http://www-ehgc.stanford.edu>
 Contact: (Dickson, Mark) md@paxil.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
 R. M.
 Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>
 Series: TRAK Plate: 192 Row: F Column: 5
 This clone was selected for full length sequencing because it
 passed the following selection criteria: Hexamer frequency ORF
 analysis, Similarity but not identity to protein.
 Location/Qualifiers
 FEATURES

source

1..1546
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="IMAGE:7111906"
/cissue_type="Testis, rat (Brown Norway)"
/clone_id="NH_MGC_237"
/lab_host="DH10B"
/note="Vector: pExpress1"
144..>1546
/codon_start=1
/product="unknown (protein for IMAGE:7111906)"
/protein_id="AAH8974.1"
/db_xref="GI:50926880"
/translation="MIMSKTAGAIPPRPSSNVRGFIRRENOVRALFHKLVIQIET
SCDPTAAVDEITGVNIGALHSQTVELKTKGIVPVAOQLHEENTQRIVEALSAS
GSDSDLSAATTKIPGLATSLGWLSFSVOLVQFKPFPIHMEHALTLRLTK
VGPEFLVLISGGCHLALVOSVDPLFLGKSIDIAPGMIDKVARLSLKIPDCST
MSGKALIEHLAKENRRFLTNPMPKANXCDSEFTGLQYDTDLTLRKKEGSIENG
OIISAADIAAVOHATAHLAKHTTHRAILFCCKNLSPRNAYLVSGVASLYLR
RALEIVANATQTLTCLPPRLCTDNGIMIANNGIERLAGIGLIHDVEDRYEKAGE
IVSWIRVALTLEDLDPSVSTHTVAHSPLNSGRGANIQSMCSCTGYMRIVRHLL
NILNEKKKKKKKKKK"

ORIGIN

Alignment Scores:

Pred. No.:	4.9e-137	Length:	1546
Score:	1725.00	Matches:	335
Percent Similarity:	91.00%	Conservative:	19
Best Local Similarity:	86.12%	Mismatches:	35
Query Match:	81.18%	Indels:	0
DB:	10	Gaps:	0

US-10-649-273-2 (1-414) x BC078974 (1-1546)

QY	I MetleutleutThlyrThAlaGlyValPhepheLyProserLyArGlyValIyr	20
Db	144 ATGCTTAAGTTAAGTGAACAGACAGACTATTCGAGAACCAACGAGATTACGTTCGT	203
QY	21 GluPhelLeuArgSerPheAsnPhetiSproGlyThrLeuPhelLeuHisIleValLeu	40
Db	204 GGATTATTAAGAAGATTTAATGTTCAACCTAGAGACACTTTTCAATCATAAATCGTGCTG	263
QY	41 GlyIleGluThrSerCysAspAspThrIlaIalavalValIaspgluThrGlyAsnval	60
Db	264 GGAATTGAAACCAGCTGTGATGACACAGCACCTGCTGTGGATGAAACCTGAAATGTG	323
QY	61 LeuGlyGluAlaIleHisSerGlnThrGluValHisIleuLySerThnglyIleValPro	80
Db	324 CTGGAGAGAAGCATGCACTCCACAGCGAAGTCCATCGAAAACAGGGGAGATTGTCCT	383
QY	81 ProAlaIaGInGlnLeuHisArgGluAsnIleGlnArgIleValIaGlnAlaIeuSer	100
Db	384 CCAGTCGCTCAACACCTTCAACAGAAATTTTCAAGCAAATATGTGAAGAAAGCTTTCT	443
QY	101 AlaSerGlyValSerProSerAspLeuSerAlaIalaIarThrIleLyProGlyLeu	120
Db	444 GCCAGTGGGGTCTCCCAAGGACCTTCAGCAATTTGACAATACATCAACACAGACTG	503
QY	121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLylys	140
Db	504 GCCCTTAACCTGGGGGTGGCTTATCCCTTTAGCCGATAGCTAGTAAACAGTTTTAAAAAG	563
QY	141 PropheIIeProIleHisIsewCInuIahIsaIaleuTrIleArgLeuThrAsnLyS	160
Db	564 CCATTATTCACCATCATCACTGAGAGCTCACGGCTGACTATTAGGCTGACCCACAA	623
QY	161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal	180
Db	624 GTCCGATTTTCTTTTAGTTCTTTTGAATTTCTGAGAGGCCACTGCTGTGGCGTTACTT	683
QY	181 GInglyValSerAspPheLeuLeuLeuGlyLySerSerLeuApriIalaIProGlyAspMet	200

Db	684	CAGGCTTTCCAGATTTTTCGCTCTCCGGGAAGTCCCTGGACATAGCCAGGCGGACATG	743
Qy	201	LeuAspIysValAlaAargIysSerLeuIleIlyshIspProGluCysSerThrMetSer	220
Db	744	CTTGACAAAGGTGGCAAGACCTTCTTTATCAAAATCCAGATATCTTACATATAGT	803
Qy	221	GIYGIYIysAlaIleGIuHISleuAlaIysGIuGIYIysAsnArgPheHISpheAspIleIys	240
Db	804	GGTGGGAAGCTATAGAACATTTGGCCAAAGAAAGAAATGATTCCACTTACTCAT	863
Qy	241	ProPheLeuHISHisAlaIysAsnCysAspPheSerPheThrGIYIleuGIHISValThr	260
Db	864	CCACCCATGACGATGCTAAGAACTGATTTTCTTTACGGGACTTCAACATGTCACC	923
Qy	261	AspIysIleIleMetIlyIysGIYIysGIuGIuGIuGIYIleGIuIysGIYIleIleuSer	280
Db	924	GATTAAGCTATTAACACACAGGAAAAAGAAAGAGGACTTGAGAAAGGGCAATCTGTCA	983
Qy	281	SerAlaIleAspIleAlaIleAlaIleThrValGIHISThrMetAlaCysHISleuValIysArg	300
Db	984	TCACCCGACACATGCTGCTCGCGGTACAGCACGCCAACAGCTGCCACTTGGCAAAAGA	1043
Qy	301	ThrHisArgAlaIleIleuPheCysIysGIuArgAspIleuProGIuAsnAsnAlaVal	320
Db	1044	ACACATGCTGTATTCTGTTTGGCACAGAAAAATTTGCTACTCCAGCTTAACGCAGTA	1103
Qy	321	IleuValIleAspGIYIYValAlaIleSerAsnPheThrIleArgAlaIleuGIuIleu	340
Db	1104	TTACTGTGTCTGGAGGTTGGCAAGTAACTCTTGACATCCGAAAGCATTTGGAAATGTGA	1163
Qy	341	ThrAsnAlaThrGluCysThrIleuIleuCysProProArgIleuCysThrAspAsnGIY	360
Db	1164	GCAAAATGCAACACATGACTTGTGTGTGTCCTCCGAGACTGTGCATGCAATAGCT	1223
Qy	361	IleMetIleAlaIlePAsnGIYIleGIuArgPheuArgAlaGIYIleuGIYIleuHISAsp	380
Db	1224	ATCATGATGTCATGGAATGGAATTGAAAGATTATGCTGTGCTTGGCATTTTACATCAT	1283
Qy	381	IleGIuGIYIleArgTYRGIuProIys	389
Db	1284	GTAGAAAGACATCCGATACGAACCAAG	1310
RESULT 12			
AR428809	LOCUS	AR428809	1526 bp DNA linear PAT 18-DEC-2003
	DEFINITION	Sequence 23 from patent US 6642041.	
	ACCESSION	AR428809	
	VERSION	AR428809.1	GI:40188595
	KEYWORDS		
	SOURCE	Unknown.	
	ORGANISM	Unknown.	
	REFERENCE	Unclassified.	
	AUTHORS	1 (bases 1 to 1526)	
	TITLE	Chen, J., Feder, J.N., Nelson, T.C., Kyrstrek, S.R. and Duclos, F.	
	JOURNAL	Polynucleotides encoding a novel metalloprotease, MP-1	
	FEATURES	Patent: US 6642041-A 23 04-NOV-2003;	
	source	location/Qualifiers	
		1..1526	
		/organism="unknown"	
		/mol_type="genomic DNA"	
ORIGIN			
Alignment Scores:			
	Pred. No.:	3,95e-108	Length: 1526
	Score:	1385.00	Matches: 267
	Percent Similarity:	100.00%	Conservative: 0
	Best Local Similarity:	100.00%	Mismatches: 0
	Query Match:	65.18%	Indels: 0
	DB:	6	Gaps: 0
US-10-649-273-2 (1-414) x AR428809 (1-1526)			
Qy	148	MetGIuAlaHISValaIleuThrIleArgIleuThrAsnIysValGIuPheProPheIleuVal	167

|||||
Db 1 ATGGAGGCTCATGACCTTACCTATAGTGTGACCAATTAAGTAATTTCTTTTAGTT 60
Qy 168 LeuLeu11eSerG1yG1yH1sCySeuLeu1a1eVal1G1nG1yVal1SerAepPheLeu 187
Db 61 CTTTGGATTTTCGGAGGTGACGTGCTGTTGGCATTAGTTCAGAGACTTTTCAGATTTTCTG 120
Qy 188 LeuLeuG1yLysSerLeuAap11eal1aProG1yAapMetLeuAap1yVal1AaArg 207
Db 121 CTTCTTGAAGAAGTCTTGGACATAGCACAGGTGAGATCTTGACAAAGTGGCAGAAAGA 180
Qy 208 LeuSerLeu11eLysH1sProG1yCySerThrMetSerG1yG1yVal1a1eG1uH1s 227
Db 181 CTTCTTAAATAAACATCCAGAGTGTCTCACATGAGTGGGAGAACCAATAGAACAT 240
Qy 228 LeuA1aLysG1nG1yAaapRphH1sPheAap11eLysProProLeuH1sH1sA1aLys 247
Db 241 TTGGCAAAACAAGAAATAGATTCTTTCATTTGACATCAACCTCCCTTGACATCATGCTAA 300
Qy 248 AapCyAapPheSerPheThrs1yLeuG1nH1sVal1ThrsAplys11e1eMetLysLys 267
Db 301 AATTGTGATTTTCTTTTACTGGACTTCAACAGCTTATGATTAATTAATGAAGAAAG 360
Qy 268 G1uLysG1uG1uG1y11eG1uLysG1yG1n1eLeuSerSer1a1aAap11eA1a 287
Db 361 GAAAAAGAGAGAGATTAAGAGAGAGGCAATCTGTCTTCAAGCAGACATTCCTGCC 420
Qy 288 ThrVal1G1nH1sThrMetA1aCyH1sLeuVal1yAArgThrsA1a1eLeuPhe 307
Db 421 ACAGTACGACACACATGAGATGCTCTTGTGAAGAAACATCGGGCTATTCGTTT 480
Qy 308 CysLysG1nAArgAapLeuLeuProG1nAaAaA1aVal1LeuVal1a1aSerG1yG1y 327
Db 481 TGTAAAGAGAGAGACTTGTACTTCAAAATATGACATCTGTTGCATCTGGTGTGTC 540
Qy 328 A1aSerAapPheTy11eArgA1aLeuG1u11eLeuThrsAa1aThrsG1nCyStr 347
Db 541 GCAAGTAACTTCTAATATCCGACAGCTCGAAATTTTAACAACGCAACACAGTGCAC 600
Qy 348 LeuLeuCyProProProAArgLeuCyStrThrsAapA1y11eMet11eA1aThrsA1nG1y 367
Db 601 TTGTGTCTCTCCCTCCAGACTATGACATTAATGATTAATGATTAATGATTAATGAT 660
Qy 368 11eG1uAArgLeuA1aG1yLeuG1y11eLeuH1sAap11eG1uG1y11eArgTyArg1u 387
Db 661 ATTGAAGAATCTACGCTGCTGGCTTGGGCAATTTACATGACATGAAGGATCCGCTATGA 720
Qy 388 ProLysCyProLeuG1yVal1aap11eSer1yG1uVal1G1yA1aSer11eLysVal 407
Db 721 CCAAAATGCTCTTGGAGTAGACATATCAAAAGAAGTTGAGAAGCTTCCATTAAGTA 780
Qy 408 ProG1nLeuLysMetG1u11e 414
Db 781 CCACAAATTAATAATGAGATA 801
RESULT 13
BX934991 1522 bp mRNA linear VMT 02-FEB-2004
LOCUS BX934991
DEFINITION Gallus gallus finished cDNA, clone CHEST189114.
ACCESSION BX934991
VERSION BX934991.1 GI:41635519
KEYWORDS
SOURCE
ORGANISM Gallus gallus (chicken)
Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 1522)
Boardman, F.E., Bonfield, J.K., Brown, W.R.A., Carder, C., Chalk, S.E.,
Croning, M.U.R., Davies, R.M., Francis, M.D., Grafton, D.V.,
Hubbard, S.J., Humphray, S.J., Hunt, P.J., Maddison, M., McLaren, S.R.,
Niblett, D., Overton, I.M., Rogers, J., Scott, C.B., Taylor, R.G.,
Tickle, C. and Wilson, S.A.

TITLE Direct Submission
JOURNAL Submitted (02-FEB-2004) Sanger Institute, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: chicken@hms.unist.ac.uk
COMMENT BBSRC/Dundee/Nottingham/Sanger/Sheffield/UMIST Gallus gallus cDNA
sequencing project
This sequence is from the
BBSRC/Dundee/Nottingham/Sanger/Sheffield/UMIST cDNA collection,
from a library constructed by Elizabeth Bosch. cDNA was prepared
from RNA extracted from muscle, normalised, and poly A-tailed.
EcoRI-NotI cut cDNA was then ligated into the vector. Vector:
pBluescript II KS(+); Site_1: EcoRI; Site_2: NotI Host: Escherichia
coli DH10B.
FEATURES
Location/Qualifiers
source
1..1522
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="layer and broiler"
/db_xref="taxon:9031"
/clone_lib="CHEST189114"
/clone_1b="CSEORBN11"
/dev_stage="adult"
ORIGIN
Alignment Scores:
Pred. No.: 3.56e-106 Length: 1522
Score: 1362.00 Matches: 258
Percent Similarity: 83.20% Conservative: 54
Best Local Similarity: 68.80% Mismatches: 63
Query Match: 64.09% Indels: 0
DB: 5 Gaps: 0
US-10-649-273-2 (1-414) x BX934991 (1-1522)
Qy 37 Lys11eVal1LeuG1y11eG1uThrsCyAapRphThrsA1a1aVal1a1aSerG1u 56
Db 254 AGACTGTGTGCGCATCGAAACACGCTGACACACGGCGGGGTGCTGAGAG 313
Qy 57 ThG1yAaA1aLeuG1yG1uA1a1eH1sSerG1nThrsG1uVal1H1sLeuLysThrsG1y 76
Db 314 GCGGGACAGGTGTGGAGAGAGCGCTGCAGAGCCAAAGAGGTCCCTCAAGACAGGT 373
Qy 77 G1y11eVal1ProProA1aG1nG1nLeuH1sAArgG1uAa11eG1nA1aG1n 96
Db 374 GGAATATTCCTCATGACAGACAGCTTCAAGAAAGCATTCCAGAAAGTATGAAG 433
Qy 97 G1uA1aLeuSerA1aSerG1yVal1SerProSerAapLeuSerA1a1eA1aThrs11e 116
Db 434 GAAGCACTCACTGCGCAGTGAATTTCTGTAATGAACCTTCTCTATTGCAACTACAGT 493
Qy 117 LysProG1yLeuA1aLeuSerLeuG1yVal1G1yLeuSerPheSerLeuG1nLeuValG1y 136
Db 494 AAACCAAGACTTCCGCTGAGCTGAGGTGAGCTGACATGACATTAAGCTGTGAC 553
Qy 137 G1nLeuLysLysProPhe11ePro11eH1sMetG1uA1aH1sA1aLeuThrs11eArg 156
Db 554 AGCTACAGAGAGCTTTCATTAACCATTCATCAATGAGGCTACAGCATTAACATGACA 613
Qy 157 LeuThrsA1nLysVal1uPheProPheLeuVal1eLeu11eSerG1yG1yH1sCySeu 176
Db 614 CTGACAGACAGATGAATATTCCTTCTTATGTTCTTACTCTCCGAGGTCACTGCATC 673
Qy 177 LeuA1aLeuVal1G1nG1yVal1SerAapRphLeuLeuG1yLysSerLeuAap11eA1a 196
Db 674 TTGGCTGACAGACAGAGATTTAGATTCTCTCTGTTGACAGTCCATAGATATAGCA 733
Qy 197 ProG1yAapMetLeuAap1yVal1AaArgLeuSerLeu11eLysH1sProG1uCy 216
Db 734 CCAAGTGAATGCTGATTAAGTACAGAGAGGCTCTTTAGTGAAGCAACCGAGATGC 793
Qy 217 SerThrsMetSerG1yG1yVal1a1eG1uH1sLeuA1aLysG1nG1yAaA1aRphH1s 236
Db 794 CACGCAATGCTGGGGGAAGGACATATGACACATGCTGCTCAACCGGAGCTGGCAAG 853

OY		260	Thraaprpysllellelmetlylsesglusgyluglugyillegluylsgylgnilleleu	279
Dd		362	ACTGATAAGCTTATATACACACAGAAGAAAAAGAAAGGCATTGAAGAGGGCAATCTCG	421
OY		280	SerSerAlaAlaAspIleAlaIalathervalgnHsttmMetAlCYeHisleuVallys	299
Dd		422	TGATCAGCTGCAGAACATTTCCTGCTCCGTACAGCATGCACAACGCCGTGCCACCTTGCGA	481
OY		300	ArgThrHisargAlaAlileuPheCybsylsglnAraspleLeuLeuproGlInasnaAna	319
Dd		482	AGAACACATCGCGCTATTCTGTTTTGGCAAGCAAGAAAATTTCCTCTCCAGCTAACCA	541
OY		320	VallLeuValAlaSerGlyGlyValAlaserasnPhetyrIleArgAlgalLeuGluile	339
Dd		542	GTAATTAAGTTGATCTCGAGAGTGCTCCAGTAACCTTGATCCGAAAGCAATTGGAAAT	601
OY		340	LeuthranAlaIAThrGlnCysThrleuLeuCysProPOPRORARleuCyethraspen	359
Dd		602	GTCCGAAATGCACACCACAGTCGACGTGTGTGTCTCCACTCCAGACCTGTCACGTGACAT	661
OY		360	GlyIleMetIleAlaITripAsnGlyIleGluarqLeuArqAlagLYleuGlyIleleuHis	379
Dd		662	GGCATCATGATTGCATTCGATGAATGGAATTTGAAGATTACGTGCTGGCGCTTTTACAT	721
OY		380	AspIlegluglyIleArgTyrgIubProLysCysProLeuGlyValAspIleserLysglu	399
Dd		722	GATGTAGAAGACATCCGATATGAACCAAATGTCTCTTTGGAGTAGACATATCCAGADA	781
OY		400	ValGlyGluAlaserIleLysValProGlnLeuLysMetGluile	414
Dd		782	GTTCGAGAGAGTCGCATTAAGTACCGCATTAATAAATGGCACCTT	826

RESULT 15

LOCUS	BX930963	1558 bp	mRNA	linear	VRT 30-MAR-2004
DEFINITION	Gallus gallus finished cDNA, clone CHEST62n16.				
ACCESSION	BX930963				
VERSION	BX930963.2	GI:4601690			
KEYWORDS					
SOURCE					
ORGANISM	Gallus gallus (chicken)				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Archosaustra; Aves; Neognathae; Galliformes; Phasianidae;				
	Phasianinae; Gallus.				
	1 (bases 1 to 1558)				
REFERENCE	Boardman,P.E., Bonfield,J.K., Brown,W.R.A., Carder,C., Chalk,S.E.,				
AUTHORS	Crofting,M.D.R., Davies,R.M., Francis,M.D., Graham,D.V.,				
	Hubbard,S.J., Humphray,S.J., Hunt,P.J., Maddison,M., McLaren,S.R.,				
	Niblett,D., Overton,I.W., Rogers,J., Scott,C.E., Taylor,R.G.,				
	Tickle,C. and Wilson,S.A.				
TITLE	Direct Submission				
JOURNAL	Submitted (29-MAR-2004) Sanger Institute, Hinxton, Cambridgeshire,				
COMMENT	CBIO ISA, UK. E-mail enquiries: chickens@hms.unist.ac.uk.				
	On Apr 1, 2004 this sequence version replaced gi:41631491.				
	BBSCR/Dundee/Nottingham/Sanger/Sheffield/UMIST Gallus gallus cDNA				
	sequencing project.				
	This sequence is from the				
	BBSCR/Dundee/Nottingham/Sanger/Sheffield/UMIST cDNA collection,				
	from a library constructed by Elizabeth Bosch. cDNA was prepared				
	from RNA extracted from limbs,				
	and poly A-trimmed. EcoRI-NotI cut cDNA was then ligated into the				
	vector. Vector: pBluescript II KS(+); Site_1: EcoRI; Site_2: NotI				
	Host: Escherichia coli DH10B.				
FEATURES	Location/Qualifiers				
source	1..1558				
	/organism="Gallus gallus"				
	/mol_type="mRNA"				
	/strain="White Leghorn, Hisex"				
	/db_xref="taxon:9031"				
	/clone="CHEST62n16"				
	/clone_lib="CSEQCHU20"				
	/dev_stage="stage 36"				

Alignment Scores:	4.57e-93	Length:	1558
Pred. No.:	1208.00	Matches:	228
Score:	82.548	Conservative:	51
Percent Similarity:	67.468	Mismatches:	59
Best Local Similarity:	56.854	Indels:	0
Query Match:	5	Gaps:	0
DB:			

US-10-649-273-2 (1-414) x BX930963 (1-1558)

QY	74	LYSTHRIgYLYIleVAlProProAlaIaIaGInGInLeuHIsARgGluAaNIleGInARg	93
DB	486	AGAGCAGGTGGAAATATATCTCTAGTGAACAGACAGCTTCAAGAAAGACATCCAGCA	545
QY	94	IlleValGInGluAlaleuSerAlaSerGlyValaSerProSerAspIleuSerAlaIleAla	113
DB	546	GTAGTAAAGGAAGCACTCACTGCTCCAGTGAAGTTTCTGTAAATGAATCTTCTGTATTTGCA	605
QY	114	ThrThrIleLYSPProGlyLeuAlaleuSerLeuGlyValaGlyLeuSerPheSerLeuGIn	133
DB	606	ACTACAGTGAACCAAGCACTTCACTGAGCCTTGAGGTGGAGCTGCAGTACAGCTTACAG	665
QY	134	LeuValaGlyGInLeuLYbLYbProPheIleProIleHIsMeGluAlaHIsAlaLeu	153
DB	666	CTGGTGAAGGGTACAGAAAGCTTTTCATACCCATTCACTACATGAGAGCTCAGCACTT	725
QY	154	ThrIleATgLeuThrAsnLYbValaGluPheProPheLeuValaLeuLeuIleSerGlyLY	173
DB	726	ACCACTCACTGACAGAGCAGCACTTAATTTCCCTTTAGTTCTTTACTCTCCGAGGT	785
QY	174	HisCYLeuLeuAlaleuValaGInGlyValaSerAspPheLeuLeuLeuGlyLYbSerLeu	193
DB	786	CACCTCATCTTGGCAGTGAAGCAAGAGAGTTTCAAGATTTCTTCTGCTTGAACAGTCCATA	845
QY	194	AspIleAlaProGlyAspMetLeuAspLYbValaIaATgArgLeuSerLeuIleLYbHIs	213
DB	846	GATATAGGACCAAGGTGACATGCTGTGAATGAAGCAAGAGCTCTTTAGTGAAGCAC	905
QY	214	ProGluCYSerThrMetSerGlyLYbValaIleGluHIsleuAlaLYbGInLYbAa	233
DB	906	CCGAGTGCACAGGATGGCTGGGGGGAAGGCAATAGACACTTGGCTCAACCGAGAC	965
QY	234	ArgPheHIsPheAspIleLYbProPheuHIsHIsAlaLYbAsnCYAspPheSerPhe	253
DB	966	TGGCAACAGTACACTTTCAGACTTCCCAAGCAAGTACGTAATGATTTTCTTTC	102
QY	254	ThrGlyLeuGInHIsValaThrAspLYbIleIleMetLYbLYbGluLYbGluGlyIle	273
DB	1026	TCCGACATTCAGAGGCTTGTCAACAAAGCATTTCTTCAAGAAAGAAAGAAAGAGTATT	108
QY	274	GluLYbGlyGInIleLeuSerSerAlaIaAspIleAlaIaThrValaGInHIsThrMet	293
DB	1086	CAGAAGGGGAATCTCTGCTCCGTTAAGACATGCTGCTGCCTGCACAGCATGATAGT	114
QY	294	AlaCYbHIsleuValaLYbArgThrHIsArgAlaIleLeuPheCYbLYbGInArgAspLeu	313
DB	1146	GCTGCTCATATATATCCAGCGGACACACGAGCAGTCTCTTGTGATGAAAACAGCAT	120
QY	314	LeuProGInaAsnAlaValaLeuValaIaSerGlyLYbValaIaAsnAspPheTYIle	333
DB	1206	TTATATACCAAAACCTGCACTCTGTTGTATATGAGAGGAGTTCAGATATACATGATATC	126
QY	334	ArgArgAlaLeuGluIleLeuThrAsnAlaThrGInCYbThrLeuLeuCYbProPro	353
DB	1286	AGAAAGGACCTGCAGACTCTGCAAAATGCAAAAGCTTTTCTTTTGTCTCTCTCA	132
QY	354	ArgLeuCYbThrAspAsnGlyIleMetIleAlaTPRAsnGlyIleGluArgLeuArgAla	373
DB	1326	AGCGCTGTCAACGAAATATGCTTATGATTCATGGAATGGCAATTCAAAGCTTGCCTGCA	138
QY	374	GlyLeuGlyIleLeuHIsAspIleGluGlyIleArgTYbGluProLYbCYbProLeuGly	393

Db	1386	GGATGTGTAATTTATACAGTACTGATGCGATCCGCTATGAACCAAGCTCCCTTGGA	1445
Qy	394	ValAspIleSerIyGluValGlyLAlaSerIleIyValProGlnLeuLys	411
Db	1446	ATTGATATTTCCAAAGAGTGAAGAGGATTCCATCAGAGTGCCAAACTAAGG	1499

Search completed: November 10, 2005, 22:36:40
 Job time : 8193.05 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: November 10, 2005, 16:48:13 ; Search time 933.3 Seconds
(without alignment)
2625.922 Million cell updates/sec

Title: US-10-649-273-2

Perfect score: 2125
Sequence: 1 MLITTTAGVFPKSKRRVY.....DISKEVGASIKVQLKMEI 414

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgr2_1/USPTO_spool_p/US10645273/runat_02112005_091337_15540/app_query.fasta_1.1429
-DB=N-geneseq_16Dec04 -QWt=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS-bits -START=1 -END=1 -MATRIX=blsum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIG=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFWT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10649273 @CGN 1 1 1063 @runat_02112005_091337_15540 -NCPU=6 -ICPR=3
-NO_MMAP -LARGEOUTRY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOC
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: N Geneseq_16Dec04:*
2: geneseqn1980s:*
3: geneseqn1990s:*
4: geneseqn2000s:*
5: geneseqn2001as:*
6: geneseqn2001bs:*
7: geneseqn2002as:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	Score	Match	Length	DB	ID	Description
No.							
1	2125	100.0	2058	6	ABA93268	ABA93268	Human O-s
2	2125	100.0	2197	6	AB576635	AB576635	DNA encod
3	2125	100.0	2572	6	ABT23207	ABT23207	Human pro
4	2088	98.3	1820	6	AAD46856	AAD46856	Human G1Y
5	2088	98.3	1820	8	ACA60887	ACA60887	Human CDN

6	2088	98.3	1821	10	AB57020	AB57020	cdna enco
7	1944	91.5	2208	10	ADA52832	ADA52832	Human cod
8	1944	91.5	2890	12	ADQ24627	ADQ24627	Human sof
9	1747	82.2	1416	6	ABX70950	ABX70950	Novel hum
10	1549	72.9	1572	6	ABQ75508	ABQ75508	Murine si
11	1385	65.2	1526	6	AB576639	AB576639	DNA encod
12	1001.5	47.1	3358	10	ADE31345	ADE31345	Human dia
13	714.5	33.6	1601	4	ABL24633	ABL24633	Drosophi
14	700.5	33.0	1385	4	AAH15110	AAH15110	Human CDN
15	687.5	32.4	3656	4	ABL24632	ABL24632	Drosophi
16	662.5	31.2	1571	4	AAC38454	AAC38454	Arabidops
17	629	29.6	1571	4	AAH12983	AAH12983	Human CDN
18	599	28.2	2734	5	AAS84622	AAS84622	DNA encod
19	531.5	25.0	4360	6	AAD48239	AAD48239	Ehrlichia
20	524	23.6	1044	8	ACA26804	ACA26804	Prokaryot
21	502	23.6	94750	12	AAFP28551	AAFP28551	Genomic f
22	502	23.6	1053	12	ADU03120	ADU03120	DNA encod
23	501	23.6	1092	6	ABQ90383	ABQ90383	M. capsul
24	500	23.5	1206	11	ABD02197	ABD02197	Pseudomon
25	499.5	23.5	1026	4	AAS54064	AAS54064	Pseudomon
26	494.5	23.3	1032	8	ACA43173	ACA43173	Prokaryot
27	492.5	23.2	1026	4	ACA42146	ACA42146	Prokaryot
28	492.5	23.2	1026	10	ADG73341	ADG73341	P aerugin
29	492.5	23.2	1059	11	ABD02280	ABD02280	Pseudomon
30	492.5	23.2	1026	10	ADG73343	ADG73343	P aerugin
31	490.5	23.1	1026	4	AAS52570	AAS52570	E. coli D
32	484.5	22.8	1014	4	ACA51431	ACA51431	Prokaryot
33	484.5	22.8	1014	8	ACA32641	ACA32641	Prokaryot
34	484.5	22.8	1014	8	ACA54107	ACA54107	Prokaryot
35	483.5	22.8	1014	8	ACA44384	ACA44384	Prokaryot
36	482.5	22.7	1074	10	ADP02056	ADP02056	Bacterial
37	482.5	22.7	1074	10	AAS56045	AAS56045	Glycoprot
38	481.5	22.7	1014	2	AAQ27645	AAQ27645	Glycoprot
39	479	22.4	1029	4	AAS53309	AAS53309	Haemophil
40	475.5	22.4	1029	8	ACA34150	ACA34150	Prokaryot
41	475.5	22.4	110000	2	AAT42063	AAT42063	Continuati
42	474.5	22.3	1032	11	ACR96255	ACR96255	Klebsiell
43	472.5	22.2	1014	10	ACR71364	ACR71364	Phototrab
44	472.5	22.2	110000	10	ACR67367	ACR67367	Continuati
45	472.5	22.2	110000	10	ACR67367	ACR67367	Continuati

ALIGNMENTS

RESULT 1	AB57020	cdna enco
AB57020	standard; cDNA; 2058 BP.	
ID	ABA93268	
XX		
AC	ABA93268;	
XX		
DT	19-APR-2002 (first entry)	
XX		
DE	Human O-sialoglycoproteinase-like protein encoding cDNA SEQ ID NO:1.	
XX		
KW	Human; O-sialoglycoproteinase-like protein; OSGP; enzyme; gene; ss.	
XX		
OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	110..1354
FT		/tag= a
FT		/product= "O-sialoglycoproteinase-like protein"
XX		
PN	CN1318550-A.	
XX		
PD	24-OCT-2001.	
XX		
PF	19-APR-2000; 2000CN-00106834.	
XX		
PR	19-APR-2000; 2000CN-00106834.	
XX		
PA	(SHAN-) SHANGHAI SHENGYUAN GENE DEV CO LTD.	
XX		

PI Mao Y, Xie Y;
XX WPI: 2002-115090/16.
DR P-PSDB; ABB05481.

PT O-sialoglycoproteinase-like protein and encoding polynucleotide, useful
for diagnosing, preventing and treating related diseases.

PS Claim 5; Page 29-30 (Disclosure); 38pp; Chinese.

XX The present sequence encodes human O-sialoglycoproteinase-like protein
CC (OSGPLP). The present invention also describes: (1) the preparation of
CC the OSGPLP protein; (2) applying the OSGPLP protein in diagnosis; (3) the
CC prevention and/or treatment of related diseases; (4) utilizing the OSGPLP
CC protein in screening its agonist, excitomotor and inhibitor and preparing
CC an antibody against the OSGPLP protein; and (5) the use of the OSGPLP
CC polynucleotide sequences, proteins, agonists, excitomotors, inhibitors
CC and antibodies in treating diseases related to the abnormal OSGPLP gene
CC and in preparing the medicine composite for the treatment

XX Sequence 2058 BP, 637 A; 400 C; 410 G; 611 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	1,94e-212	Length:	2058
Score:	2125.00	Matches:	414
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-10-649-273-2 (1-414) x ABA93268 (1-2058)

QY	1	MetLeuIleLeuThrLysThrAlaGlyValPhePheLysProSerLysArgLysValIyr	20
DB	110	ATGCTAATCTTCACTAAGAGCTCAGAGAGTTTAAAAACCAATCAAAAGAAAGCTTAT	169
QY	21	GIuPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisLysIleValIleu	40
DB	170	GAAATTTTAAAGAAAGTTTAAATTTTCATCTGGAACCTATTTCTATATAAAATGATTTG	229
QY	41	GLYIleGIuThrSerCysAspAspThrAlaAlaValAspGIuThrGIuAsnVal	60
DB	230	GGAATTTGAAACTAGTTTGATGATATACAGAGCTCGTGTGTGATGAAACTGGAAATGTG	289
QY	61	LeuGIuValAlaIleHisSerGIuThrGIuValHisLeuLysThrGIuLysIleValPro	80
DB	290	TTGGAGAAAGCAATACATTTCCAAACTGAAGTTCAATTTAAAAACAGTGGATTTGTTCT	349
QY	81	ProAlaAlaGIuThrLeuHisArgGIuValenIleGIuArgIleValGIuThrAlaLeuSer	100
DB	350	CCAGCAGCTCAACACCTTCACAGAGAAATATTTCAACGAATATACAGAAAGCTCTTTCT	409
QY	101	AlaSerGIuValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGIuLysLeu	120
DB	410	GCCAGTGGAGTCTCTCCAAAGTACCTCTCAGCAATTCGAACTTCACTTAAAAACAGACTT	469
QY	121	AlaLeuSerLeuGIuValGIuLysSerPheSerLeuGIuLeuValGIuLysLeuLysIys	140
DB	470	GCTTTAAGCTCGAGTGGCTTATCATTTAGCTTACAGTGTGTAGAGAGTTAAAAAAG	529
QY	141	ProPheIleProIleHisIleMetGIuAlaHisAlaLeuThrIleArgLeuThrAsnLys	160
DB	530	CCATTCATTTCCATTCATCATATGAGGCTCAATGACCTTATTTAGTTGACCAATTA	589
QY	161	ValGIuPheProPheLeuValLeuLeuIleSerGIuLysIleCysLeuLeuAlaLeuVal	180
DB	590	GTAAGATTTCTTTTATTTTATTTCTTTTATTTCTTGAGAGTACATGTCGTGGCAATTA	649
QY	181	GIuGIuValSerAspPheLeuLeuLeuGIuLysSerLeuAspIleAlaProGIuAspMet	200
DB	650	CAAGAGTTTCAATTTTCTGCTTCTTGAGAAAGTCTTTGACATAGCACCAGGTGACATG	709
QY	201	LeuAspLysValAlaArgArgLeuSerLeuIleLysIleProGIuLysSerThrMetSer	220

DB	710	CTTGACAAAGTGGCAAGAGACTTTCTTTAATAAAACATCCAGAGCTCCACATGAGT	769
QY	221	GIuGIuLysValAlaIleGIuHisLeuAlaLysGIuGIuLysAsnArgPheHisPheAspIleLys	240
DB	770	GGTGGGAAAGCCATTAACATTTTGGCCAAACAAAGAAATAGATTTTCAATTTTGACATCA	829
QY	241	ProProLeuHisIleHisAlaLysAsnCysAspPheSerPheThrGIuLysGIuHisValThr	260
DB	830	CTTCCTTCATCATCTCTTAAAAATTTGATTTTCTTTTAACTGAGCTTCAACAGCTTACT	889
QY	261	AspLysIleIleLeuLysLysGIuLysGIuLysGIuLysIleGIuLysGIuLysIleLeuSer	280
DB	890	GATTAATATATATGAAAAAGGAAAAAGGAAAGTATTTGAGAAAGGCAATCTCTCT	949
QY	281	SerAlaAlaAspIleAlaThrValGIuHisIleThrMetAlaCysHisIleuValLysArg	300
DB	950	TCAGCAGCAGACATTCCTCCACAGATACAGACACAAATGCAATGCTATCTTTGAAAGA	1009
QY	301	ThrHisArgAlaIleLeuPheCysLysGIuArgAspLeuLeuProGIuAsnAlaVal	320
DB	1010	ACACATCGGGCTATTTCTGTTTGTATAGCAGAGACTGTACTCAAAATATATGCGTA	1069
QY	321	LeuValAlaSerGIuLysValAlaLysAsnPheThrIleArgArgAlaLeuGIuLysLeu	340
DB	1070	CTGTTTGACATCTGTGTGTGTGCGCAAGTAACTTCTATATCCGAGAGCTTGAAATTTTA	1129
QY	341	ThrAsnAlaThrGIuLysThrLeuLysCysProProArgArgLeuCysThrAspAsnGIu	360
DB	1130	ACAAACGCAACACAGGCACTTTGTGTGTGCTCTCCACGACTATGCACTGATTAATGCG	1189
QY	361	IleMetIleAlaThrAsnGIuLysIleGIuArgLeuArgAlaGIuLysLeuHisAsp	380
DB	1190	ATTATATATTCATGAAATGATTTGAAAGCTACAGCTGCTTGCGCATTTTACATGAC	1249
QY	381	IleGIuGIuLysIleArgArgGIuProLysCysProLeuGIuValAspIleSerLysGIuVal	400
DB	1250	ATGGAAGCATTCGCTATGATACCAAAATGTCTCTTGGAATGACATATCAAAAGAGTT	1309
QY	401	GIuGIuLysSerIleLysValProGIuLysLeuMetGIuLysIle	414
DB	1310	GGAGAAAGCTTCATTAATAAGTACCAATTAATAATGAGATA	1351

RESULT 2
ABS76635 standard; DNA; 2197 BP.
ID ABS76635;
AC ABS76635;
XX
DT 11-DEC-2002 (first entry)
XX
DE DNA encoding novel human metalloproteinase MPI.
XX
XX Metalloproteinase; MP-1; immune disorder; glutamate transport; cancer;
KW motor neuron disorder; amyotrophic lateral sclerosis; ALS; diabetes;
KW reproductive disorder; Klinefelter's syndrome; germinal cell aplasia;
KW genital wart; metabolic disorder; premature puberty; Kallman syndrome;
KW Cushing's syndrome; neurodegenerative disease; Alzheimer's disease;
KW Parkinson's disease; Huntington's disease; Tourette syndrome; sepsis;
KW liver disease; renal disease; immune disorder; rheumatoid arthritis;
KW acquired immunodeficiency syndrome; AIDS; pulmonary disease; pneumonia;
KW emphysema; cystic fibrosis; vascular disorder; inflammatory disorder;
KW neurological disorder; gene; ds.
XX
OS Homo sapiens.
XX
PN W0200272751-A2.
XX
XX 19-SEP-2002.
PD
XX 05-FEB-2002; 2002WO-US003353.
PF
XX 05-FEB-2001; 2001US-0266518P.
PR

PR 10-APR-2001; 2001US-0282814P.
XX (BRIM) BRISTOL-MYERS SQUIBB CO.
XX Chen J, Feder J, Nelson TC, Duclos F, Krystek S;
XX WPI; 2002-723329/78.
XX P-PSDB; ABG96478.
XX
XX New isolated nucleic acid encoding MP-1 protein, useful for preventing,
XX treating, or ameliorating diseases associated with aberrant
XX metalloproteinase activity, e.g. immune, metabolic, inflammatory and
XX neurological disorders.
XX
XX Claim 1; Fig 1A-C; 473pp; English.
XX
XX The invention describes an isolated nucleic acid molecule (I) encoding a
XX metalloproteinase (MP-1). (I) is useful for preventing, treating, or
XX ameliorating a medical condition, particularly an immune disorder, an
XX aberrant glutamate transport or motor neuron disorder, such as
XX amyotrophic lateral sclerosis (ALS), its juvenile form or an ALS-like
XX condition. The compositions and methods are also useful for diagnosing,
XX prognosticating, treating, ameliorating and/or treating disorders
XX associated with MP-1 activity, e.g. diabetes, cancer, reproductive
XX disorders (e.g. Klinefelter's syndrome, genital warts, or germinal cell
XX aplasia), metabolic disorders (e.g. premature puberty, Kallman syndrome,
XX or Cushing's syndrome), neurodegenerative diseases (Alzheimer's disease,
XX Parkinson's disease, Huntington's disease or Tourette syndrome), liver
XX and renal diseases and immune disorders (e.g. AIDS, rheumatoid arthritis
XX or sepsis), pulmonary diseases (e.g. pneumonia, emphysema or cystic
XX fibrosis) and vascular, inflammatory and neurological disorders (e.g.
XX Alzheimer's disease or Parkinson's disease). This sequence represents a
XX metalloproteinase MP1 polynucleotide
XX
XX
SQ Sequence 2197 BP; 681 A; 441 C; 439 G; 636 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 2,146-212 Length: 2197
Score: 2125.00 Matches: 414
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: Gaps: 0
US-10-649-273-2 (1-414) x ABS76635 (1-2197)

QY 1 MetLeuIleLeuThrLysThrAlaGlyValPhePheIysProSerLysArgLysValIyr 20
DB 231 ATGCTAAATCTTGACCTAGACCTGACAGAGCTTTTAAACCATCAAAAGAAAGTTTAT 290
QY 21 GluPheLeuArgSerPheAspPheHisProGlyThrLeuPheLeuHisLysValIleu 40
DB 291 GAATTTTAAAGAGATTAAATTTTCACTGGAACAATATTTCTCATTAATAATGATATG 350
QY 41 GlyIleGluThrSerCysAspAspThrAlaAlaValAspGluThrGlyAsnVal 60
DB 351 GGAATTTGAATCTAGTTGATGATATACAGACGCTGCTGTGTGATGAATGAACTGAAATG 410
QY 61 LeuGlyGluAlaIleHisSerGlnThrGluValHisLeuLysThrGlyValIleValPro 80
DB 411 TTGGAGAGAGCAATACATCTCCAACTGAACTTATTTAAACAGAGTGGATGTTCT 470
QY 81 ProHlaaIaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100
DB 471 CCACAGCTCAACACCTTCAACAGAAATATTTCAACGATGTCACAGAACCTCTTCT 530
QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeu 120
DB 531 GCCAGTGAAGTCTCTCCAGAGACCTCTCAGCAATTTGCACTTAAACCAAGACACTT 590
QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLys 140
DB 591 GCTTTAAGCCTGGAGGTGGGCTTATCATTTAGCTTACAGCTGTGTAGAGACGTTAAAAAG 650

QY 141 ProPheIleProIleHisIsmecGluAlaHisAlaLeuThrIleArgLeuThrAsnLys 160
DB 651 CCATTCAATCCCATTCATCATATGAGAGCTCATGCACTACTATTAAGTGGACCAATATA 710
QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180
DB 711 GTAGAAATTTCTTTTATGATCTTTGATATTTCTGAGAGTCACTGTCTGGTGGCATTTAGT 770
QY 181 GlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet 200
DB 771 CAAGAGCTTCAAGTTTCTGCTTCTTGAAAGCTTTGGACATAGCAACAGGAGCACTG 830
QY 201 LeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGluCysSerThrMetSer 220
DB 831 CTGACAGAGTGGCAAGAAAGACTTTCTTTAATAAATCATCCAGAGTCTCCACATAGT 890
QY 221 GlyGlyLysValAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLys 240
DB 891 GGTGGGAAAGCCATAGAACATTTGGCCAAACAGAAATAGATTTCATTTTGACATCAAA 950
QY 241 ProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260
DB 951 CTCCCTTGATCATGCTAAATAATGTGATTTTCTTTACTGAGACTTCAACAGCTTACT 1010
QY 261 AspLysIleIleMetLysLysGlyLysGlyGlyIleGlyLysGlyGlnIleLeuSer 280
DB 1011 GATTAATATATATATGAAAAAGAAAAAGAAAGAGTATTGAGAAAGGCAAACTCTCT 1070
QY 281 SerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCysHisLeuValLysArg 300
DB 1071 TCAGCAGACAGACATTCGTCGCACAGTACACACATGCAATGCAATCTTGTGAAAGA 1130
QY 301 ThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuProGlnAsnAlaVal 320
DB 1131 ACAATCGGGCTATTTCTGTTTGTAAAGACAGAGACTGTGTACTCAAAATTAATGCGTA 1190
QY 321 LeuValAlaSerGlyGlyValAlaSerAspPheThrIleArgArgAlaLeuGluIleLeu 340
DB 1191 CTGGTTGATCTGTGTGTGTCGCAAGTAACTTATATCCGCAAGCTCTCGAAATTTTA 1250
QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeuCysThrAspAsnGly 360
DB 1251 ACAAAAGCAACACAGTCACTTTGTTGTGTCCTCTCCAGACTATGCACTGATTAATGCG 1310
QY 361 IleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380
DB 1311 ATTATGATTGCAATGGAATGTATTTGAAAGCTACGCTGCTGGCATTTTATCAATGAC 1370
QY 381 IleGluGlyIleArgTrpGluProLysCysProLeuGlyValAspIleSerLysGluVal 400
DB 1371 ATTAGAAGGCATTCGGCTATGAAACCAAAATGCTCTTGAGATAGACATATCAAAAGAGTT 1430
QY 401 GlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 414
DB 1431 GGAAGAGCTTCCATTAAGTACCAATTAATAATGAGATA 1472

RESULT 3
ABT23207
ID ABT23207 standard; DNA; 2572 BP.
XX
XX ABT23207;
XX
XX 01-MAY-2003 (first entry)
XX
XX Human protein modification + maintenance molecule DNA SEQ ID No 36.
XX
XX Cytostatic; antiarteriosclerotic; anti-HIV; antiallergic; anticonvulsant;
XX cerebroprotective; antiparkinsonian; nootropic; antiinflammatory;
XX antitumor; hepatotropic; gynecological; antibacterial; virucide;
XX protozoicide; antiparasitic; cell proliferative disease; PMOD;
XX protein modification and maintenance molecule; immunogenic fragment;
XX cancer; autoimmune; inflammatory disease; neurological disorder;

Db	1154	ACAAACGACACAGTGCATCTTGTGTGCTCTCTCCACACATACATGCATGATTAATGCG	1223
Oy	361	llelellleallatpaspnglylleleluarigleuargalaglyleuglylleleuhisasp	380
Db	1224	ATTATGATTCATGCATGATATGATTAAGACCTACGTCGTGGCTTGGCATTTTACATGAC	1283
Oy	381	lleleluaglylleargtyrgluprolyecysproleuglyvalasppilleseerlysgluval	400
Db	1284	ATGAGAGGCATCTCCGCTATGAACCAAAAAGTCCTCTTGAAGTAGACATATCAAAAAGAACTT	1343
Oy	401	glyglualaaserilleylevalprogluineuylsmeuile	414
Db	1344	GGAGAGCTTCATTAAGTACACACATTAATAATGGAGATA	1385
RESULT 4			
AAD46856	ID	AAD46856 standard; cDNA; 1820 BP.	
XX	AC	AAD46856;	
XX	DT	27-JAN-2003 (first entry)	
XX	DE	Human glycoprotease 28472 cDNA.	
KW	XX	Human; adenosine deaminase; seven transmembrane domain receptor; cancer;	
KW	XX	7TM; glycoprotease; immune disorder; Iga deficiency; allergy; arhythmia;	
KW	XX	rheumatoid arthritis; diabetes; atherosclerosis; cardiovascular disorder;	
KW	XX	hypertension; ischaemic heart disease; obesity; myocardial infarction;	
KW	XX	endothelial cell disorder; Grave's disease; psoriasis; brain disorder;	
KW	XX	Parkinson's disease; Alzheimer's disease; haematopoietic disorder;	
KW	XX	cerebral oedema; metabolic disorder; liver disorder; platelet disorder;	
KW	XX	chromosome mapping; tissue typing; gene therapy; neuroprotective;	
KW	XX	cystostatic; anorectic; cardiant; haemostatic; gene; ss.	
OS	XX	Homo sapiens.	
XX	XX		
FH	XX	Key	Location/Qualifiers
FT	XX	CDS	146..1390
FT	XX		/*tag= a
FT	XX		/product= "Human 28472 protein"
FT	XX		/note= "This region is specifically claimed as SEQ ID NO: 6 in claim 1 of the specification"
FT	XX		
XX	XX	WO200274960-A2.	
PN	XX		
PD	XX	26-SEP-2002.	
XX	XX		
PF	XX	08-NOV-2001; 2001WO-US051427.	
XX	XX		
PR	XX	08-NOV-2000; 2000US-0246768P.	
PR	XX	08-NOV-2000; 2000US-0246772P.	
XX	XX	15-NOV-2000; 2000US-0249185P.	
XX	XX		
PA	XX	(MILL-) MILLENNIUM PHARM INC.	
PI	XX		
PI	XX	Leidy KR, Kapeller-Libermann R, Glucksmann M;	
XX	XX		
DR	XX	MPi; 2002-759898/82.	
DR	XX	P-PEDB; AAE29234.	
XX	XX		
PT	XX	New 38650, 28472, 5495, 65507, 81588 or 14354 nucleic acid molecules,	
PT	XX	useful for diagnosing and/or differentiation (e.g. colon or	
PT	XX	hematopoietic, brain, pain, metabolic, liver or platelet disorders, and	
PT	XX	in pharmacogenomics.	
XX	XX		
PS	XX	Claim 1; Fig 8; 178pp; English.	
XX	XX		
CC	CC	The present invention relates to novel 38650, 28472, 5495, 65507, 81588	
CC	CC	or 14354 nucleic acid molecules which encode adenosine deaminase, glyco-	
CC	CC	protease or seven transmembrane domain (7TM) receptor family members.	
CC	CC	Sequences of the invention are useful in diagnosing and treating cancer	
CC	CC	or aberrant cellular proliferation and/or differentiation (e.g. colon or	
CC	CC	lung cancer), immune disorders (e.g. selective Iga deficiency, rheumatoid	

CC	arthritis, diabetes, allergy), heart or cardiovascular disorders (e.g.,
CC	hypertension, atherosclerosis), arrhythmias, ischaemic heart disease,
CC	myocardial infarction, thrombus) including endothelial cell disorders
CC	(e.g. psoriasis, Grave's disease), haematospicetic disorders, brain
CC	disorders (e.g. cerebral oedema, Parkinson's or Alzheimer's disease),
CC	pain and metabolic disorders (e.g. obesity), liver disorders or platelet
CC	disorders. They are also useful in screening assays, predictive medicine
CC	(e.g. diagnostic assays, prognostic assays, monitoring clinical trials
CC	and pharmacogenetics) and prophylactic and therapeutic methods. The
CC	nucleic acids may also be used in chromosome mapping, tissue typing and
CC	forensic biology and as surrogate markers. Sequences of the invention are
CC	also used in gene therapy. The present sequence is human glycoprotease
CC	28472 cDNA
SX	
XX	Sequence 1820 BP; 543 A; 365 C; 393 G; 518 T; 0 U; 1 Other;
Alignment Scores:	
Pred. No.:	1,24e-208 Length: 1820
Score:	2088.00 Matches: 407
Percent Similarity:	99.03% Conservative: 3
Best Local Similarity:	98.31% Mismatches: 4
Query Match:	98.26% Indels: 0
DB:	Gaps: 0
US-10-649-273-2 (1-414) x AAD46856 (1-1820)	
OY	1 MetLeuIleLeuThrTySThraLagIyValPhePhelyProSerLySarIgyValTyr 20
Db	146 ATGCTATCTTGACTGAAGCTCAGAGAGTTTTTTTAACCATCAAAAGAAGTTTAT 205
OY	21 GluPheLeuArgSerPheAsnDheHisSPROGIYHrLeuPheLeuHisLysIleValLeu 40
Db	206 GAATTTTTAAAGATTTCATTTAATTTTCATCTCGAMACATATTTCTTCAATAAATGTATTC 265
OY	41 GlyTlLeuInHserCySaSpaSprhTrAlaaIaValValAspGluInThGlyVaSPVal 60
Db	266 GGAATTGAAACTAGTGGTAGATATVACAGACGCTGCGTGGAGTGAAGAACTGGAAGTGC 325
OY	61 LeuGlYgLuAlalIehIsSerGlnThrgIuValHisLeuLysThngIyglYlIeValPfo 80
Db	326 TTGGGAGAACCATATACATTCGCCAACTGAAGTTCATTAAAAACAAGTGGATTTGTTCT 385
OY	81 ProAlaAlaGlnGlnLeuNhiIsArgIuAsnIlEgInatgIIeValGlnGluAlaLeuSer 100
Db	386 CCAGCAGCTCAACAGCTTCACAGAGAAATATTCACACAAATAGTACAGAAGCTCTTCT 445
OY	101 AlISerGlyValSerProSeraspLeuSerAlalealattPrTrllIeLysPRoGlyLeu 120
Db	446 GCCAGTGAAGTCTCTCCAAGTACCCTTCAGCAATGGACATCACATMAAACAGAGACTT 505
OY	121 AlaLeuSerLeuGlYvalGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLysLys 140
Db	506 GCITTAAGCCTGGAGATGGGCTTATCATTTAGCTTACACGCTGTAGACACTTAAAAAG 565
OY	141 ProPheIleProIleHisIsmetGluAlahisalaLeuThrllIeargleuThrasnLys 160
Db	566 CCAATCATTCCTCCATTCATCATATGAGAGGCTCATGCACTTACATTAAGGTGACCAATAAA 625
OY	161 ValGluPheProPheLeuValLeuLeuLieserGlyglYhiIsCybleuLeuAlaLeuVal 180
Db	626 GTAGAAATTCCTTTTTCAGTTCTTTTGAGATTCCTGAGAGTCACTGCTGTGGCAATGATT 685
OY	181 GIAnGIYValSeraspPheLeuLeuLeuGlyLysSerLeuAspIlealaproGlyVaSPMet 200
Db	686 CAAGAGATTTCAATTTTCTGCTTCTTGGAAAGTCTTTGGACATRGCACAGGTGAACATG 745
OY	201 LeuaspLysValAlaargIrgLeuSerLeuIleLysHisiproGluCYeSerThmeSer 220
Db	746 CTTGACAAAGGTGCAGAGAAGACTTTCTTTAATPAACAATCCAGAGTGTCCACCATAGAT 805
OY	221 GlyGLYLYValAlIlegluHisLeuAlalySGlngIyaMaTqPhehisPheAspIleLys 240
Db	806 GGtGGGAAGCCCTATMACATTTGGCAAACAGAAATAGATTCAATTTGACATCAAA 865

QY 241 ProProLeuHshHshAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260
 DB CCTCCCTTGCACTGATCTGTAATAATGTAATTTTCTTTACTGAGACTTCAACCGTTACT 925
 QY 261 AspLysIleIleMetLysGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 280
 DB GATAAATAATATGAAACACAGAAACAGAAAGATTAAGAAAGGGCAAAATCCGTCT 985
 QY 281 SerAlaAlaAspIleAlaIleAlaIleAlaIleAlaIleAlaIleAlaIleAlaIleAla 300
 DB TCAGCAGCAGCAATGCTGCCACAGTACGACACACATGCGCATGCTCTTGGAAGA 1045
 QY 301 ThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuProGlnAsnAlaVal 320
 DB ACACATCGGGCTATCTGTTGTTGTAAGCAGAGACTTTTACTCAAAATATGACAGTA 1105
 QY 321 LeuValAlaSerGlyGlyValAlaAspAsnAspPheTyrIleArgArgAlaLeuGlnIleLeu 340
 DB CTGGTTGATCTGGTGGTGTCCAGTACTTATATCCGACAGCTCTGGAAATTTTA 1165
 QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeuLeuCysThrAspAsnGly 360
 DB AACAAACGCAACACAGTGCCTTGTGTCTCTCCCAAGCTATGACCTGATATATGAC 1225
 QY 361 IleMetIleAlaIleAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380
 DB ATTATGATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1285
 QY 381 IleGluGlyIleArgTyrGlyLeuProLysCysProLeuGlyValAspIleSerLysGlyVal 400
 DB ATGAGAGGATCGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1345
 QY 401 GlyGluAlaSerIleLysValProGlnLeuLysMetGlnIle 414
 DB 1346 GGAGAGAGCTTCATTAAGTACCAACATTAATAATGAGATA 1387
 DB
 RESULT 5
 ACA60887
 ID ACA60887 standard; cDNA; 1820 BP.
 XX
 AC ACA60887;
 XX
 DT 08-JUL-2003 (first entry)
 XX
 DE Human cDNA 28472 encoding a glycoprotease.
 XX
 KW Human; ss; gene; cancer; aberrant cellular proliferation;
 KW differentiation; immune disorders; heart disorder; brain disorder;
 KW cardiovascular disorder; endothelial cell disorder; pain disorder;
 KW hematopoietic disorder; blood vessel disorder; metabolic disorder;
 KW liver disorder; platelet disorder; glycoprotease.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 146..1390
 FT /tag= a
 FT /product= "Glycoprotease"
 FT /note= "This CDS is specifically claimed in claim 1"
 XX
 PN US2003009017-A1.
 XX
 PD 09-JAN-2003.
 XX
 PF 08-NOV-2001; 2001US-00012140.
 XX
 PR 08-NOV-2000; 2000US-0246768P.
 PR 08-NOV-2000; 2000US-0246772P.
 PR 15-NOV-2000; 2000US-0249185P.
 XX
 PA (LEIBY/) LEIBY K R.
 PA (KAP/) KAPPELLER-LIBERMANN R.

PA (GLUC/) GLUCKSMANN M A.
 XX
 PI Leiby KR, Kapeller-libermann R, Glucksmann MA;
 XX
 DR WPI: 2003-428888/40.
 DR P-PSDB; AB009569.
 PT
 PT New isolated 38650, 28472, 5495, 65507, 81588 and 14354 nucleic acid
 PT molecules, useful for diagnosing, treating cancer, pain, or immune,
 PT heart, endothelial cell, hematopoietic, blood vessel, brain, metabolic
 PT and liver disorders.
 XX
 PS Claim 2; Fig 8; 90pp; English.
 XX
 CC The invention relates to an isolated 38650 (encoding adenosine
 CC deaminase), 28472 (encoding a glycoprotease), 5495 (encoding a 7
 CC transmembrane receptor), 65507, 81588 or 14354 nucleic acid molecule (or
 CC a sequence which is at least 60% identical to the six nucleic acids or
 CC their open reading frames, fragments of at least 15 nucleotides,
 CC naturally occurring variants, or a DNA insert of the plasmid deposited
 CC with the American Type Culture Collection as Accession No. not defined in
 CC the specification, which encodes the amino acid sequence). Also included
 CC are a host cell containing the nucleic acids (used to produce the
 CC proteins), the encoded proteins, an antibody that selectively binds to
 CC the polypeptide, and identifying a compound that binds to/modulates the
 CC activity of the polypeptide. The nucleic acid molecules, polypeptides and
 CC methods are useful for diagnosing, treating cancer, aberrant cellular
 CC proliferation and/or differentiation, immune disorders, heart disorders,
 CC cardiovascular disorders including endothelial cell disorders, pain
 CC and metabolic disorders, liver disorders and platelet disorders (many
 CC examples of these disorders are given in the specification). The present
 CC sequence is the Human cDNA 28472 encoding a glycoprotease
 XX
 SQ Sequence 1820 BP; 543 A; 365 C; 393 G; 518 T; 0 U; 1 Other;
 XX
 Alignment Scores:
 Pred. No.: 1,24e-208 Length: 1820
 Score: 2088.00 Matches: 407
 Percent Similarity: 99.03% Conservative: 3
 Best Local Similarity: 98.31% Mismatches: 4
 Query Match: 98.26% Indels: 0
 DB: 8 Gaps: 0
 US-10-649-273-2 (1-414) x ACA60887 (1-1820)
 QY 1 MetLeuIleLeuThrLysThrAlaGlyValPhePheLysProSerLysValGlyValTyr 20
 DB ATGCTAATCTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 205
 QY 21 GlnPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisLysIleValLeu 40
 DB GAAATTTTAAAGAGTTTAAATTTTCACTCGAAGCACTAATTTCTTCAATTAATAGTATG 265
 QY 41 GlyIleGluThrSerCysAspAspThrAlaAlaValAspGluThrGlyAsnVal 60
 DB GGAATTCAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 325
 QY 61 LeuGlyGluAlaIleHisSerGlnThrGluValHisLeuLysThrGlyIleValPro 80
 DB TTGGGAGAAACAAATATCTCCAAATGAAAGTTCAATTAATAACAGGTGGGATTTGTTCT 385
 QY 81 ProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGlnAlaLeuSer 100
 DB CCAGCAGCTCAACAGCTTCAAGAGAAATATTCAGAAATAGTACAAAGAGCTCTTTCT 445
 QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeu 120
 DB GCCAGTGGAGTCTCTCAAGTGAAGCTTCCAGCAATTCGAATACATTAACACGAGACTT 505
 QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLysLys 140
 DB GCTTTAAGCTCGGAGTGGGCTTATCATTTAGCTTACAGCTGATGAGACAGTTAAAGAG 565

QY 141 ProPheleProIleHisMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLys 160
 Db 566 CCATTCATTTCCCATCATCATATATGAGGCTCATGCACTTAATATGTTGATGACCAATATA 625
 QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisGlyLeuLeuAlaLeuVal 180
 Db 626 GTAGAAATTCCTTTTATGTTCTTTTGAATTTCTGGAGGTCACGTGCTGTTGGACATTAAGTT 665
 QY 181 GlnglyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet 200
 Db 686 CAAGAGATTTCAGATTTTCTGCTTCTTGAAAGTCTTTGGACATACACAGGTGACATG 745
 QY 201 LeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGluCysSerThrMetSer 220
 Db 746 CTTCAGAGTGGCAAGAAAGCTTCTTAAATAAACATCCAGAGTGTCCACACATGAGT 805
 QY 221 GlyGlyLysAlaIleGluHisIleuAlaLysGlnGlyLysAsnArgPheHisPheAspIleLys 240
 Db 806 GGTGGAAAGCCATAGAAACATTGGCCAAACAAAGAAATAGATTTCATTTTGACATCAAA 865
 QY 241 ProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260
 Db 866 CTTCCCTTGCAATCATGCTTAATAATTTGTGATTTTCTTTTACTGCACTTCAACACGTTACT 925
 QY 261 AspLysIleIleMetLysLysGluLysGluGlyIleGluLysGlyGlnIleLeuSer 280
 Db 926 GATATAAATATATGAAACACGAAACAGAAAGAAAGATATTGAGAAAGGCGCAATTCGTCT 985
 QY 281 SerAlaAlaAspIleAlaIleAlaThrValGlnHisIleMetAlaCysHisIleuValLysArg 300
 Db 986 TCAGACAGACGATTTGCTGCCACAGTACAGCACACATATGTCATCTTGTGAAATAA 1045
 QY 301 ThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuProGlnAsnAlaVal 320
 Db 1046 AACAATCGGGCTATTCTGTTTGTAAAGCAGAGAGACTTGTACTTCAATAATATGACATA 1105
 QY 321 LeuValAlaSerGlyGlyValAlaSerAsnPheIleArgArgAlaLeuGluIleLeu 340
 Db 1106 CTGCTGCATCTCGTGCGTGTCCAGATACCTTATATCCGACAGCTCTGAAATTTTA 1165
 QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeuCysThrAspAsnGly 360
 Db 1166 ACAAAACGACACACATGCACTTGTGTGTCTCTCCACAGACTATGACCTGATATATGCG 1225
 QY 361 IleMetIleAlaIleAsnGlyIleGluArgLeuAlaGlyLeuGlyIleLeuHisAsp 380
 Db 1226 ATTATGATTGCAATGGAATGTAATTGAAACACGCTGCGCTTGGCATTTTACATGAC 1285
 QY 381 IleGluGlyIleArgGlyLeuProLysCysProLeuGlyValAspIleSerLysGluVal 400
 Db 1286 ATAGAAAGCATCCGCTATGAAACCAAAATGTCCTCTTGAGTACATATCAAAAGAAAGTT 1345
 QY 401 GlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 414
 Db 1346 GGAGAGCTTCATATAAGTACACACATTTAAATGAGATATA 1387
 RESULT 6
 ABSS7020
 ID ABSS7020 standard; cDNA; 1821 BP.
 XX
 AC ABSS7020;
 XX
 DT 30-JAN-2003 (first entry)
 XX
 DE cDNA encoding novel human glycoprotease 28472.
 XX
 KM Cancers; aberrant cell proliferation; aberrant cell differentiation;
 KM breast cancer; ovarian cancer; prostate cancer; colon cancer;
 KM lung cancer; immune disorder; heart disorder; cardiovascular disorder;
 KM endothelial disorder; hematopoietic disorder; blood vessel disorder;
 KM brain disorder; pain; metabolic disorder; liver disorder; diabetes;
 KM platelet disorder; carcinoma; sarcoma; leukemia; Hodgkin's disease;

KM autoimmune disorder; hypertension; atherosclerosis; heart failure;
 KM myocardial infarction; ischaemic heart disease; Crohn's disease;
 KM Kawasaki's disease; Kawasaki syndrome; Raynaud's disease; aneurysm;
 KM cerebral ischaemia; peripheral neuropathy; Alzheimer's disease;
 KM Parkinson's disease; anorexia nervosa; cachexia; glycoprotease; gene; ss.
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 147..1391
 FT /tag= a
 FT /product= "Glycoprotease 28472"
 FT /note= "Specifically claimed in claim 1"
 PN MO200277233-A2.
 XX
 PD 03-OCT-2002.
 XX
 PF 08-NOV-2001; 2001MO-US046724.
 XX
 PF 08-NOV-2000; 2000US-0246768P.
 PR 08-NOV-2000; 2000US-0246772P.
 PR 15-NOV-2000; 2000US-0249185P.
 XX
 PA (MILL-) MILLENNIUM PHARM INC.
 PI
 PI Leiby KR, Kapeller-Libermann R, Glucksmann M;
 DR WPI; 2003-029938/02.
 DR P-PSDB; ABG71162.
 XX
 XX New adenosine deaminase, glycoprotease and seven transmembrane domain
 PT nucleic acids and polypeptides, designated 38650, 28472, 5495, 65507,
 PT 81588 and 14354, useful for treating e.g. leukemias, Hodgkin's disease or
 PT hypertension.
 XX
 PS Claim 2; Fig 8A-B; 178pp; English.
 XX
 CC The invention describes isolated 38650, 28472, 5495, 65507, 81588 and
 CC 14354 nucleic acid molecules (I) and their encoded polypeptides (II). The
 CC 38650 nucleic acid molecule comprises a sequence encoding adenosine
 CC deaminase. The 28472 nucleic acid molecule comprises a sequence encoding
 CC a human glycoprotease. The 5495, 65507, 81588 and 14354 comprise
 CC sequences that encode a human seven transmembrane domain (7TM). The
 CC 38650, 28472, 5495, 65507, 81588 or 14354 nucleic acid and polypeptide
 CC sequences are useful for diagnosing, preventing or treating a subject
 CC with or at risk of developing a disorder, e.g. cancer or aberrant
 CC cellular proliferation and/or differentiation (e.g. breast, ovarian,
 CC prostate, colon or lung cancer), immune disorders, heart disorders,
 CC cardiovascular disorders, endothelial disorders, hematopoietic disorders,
 CC blood vessel disorders, brain disorders, pain and metabolic disorders,
 CC liver disorders or platelet disorders. These disorders include carcinoma,
 CC sarcoma, leukaemia, Hodgkin's disease, autoimmune disorders,
 CC hypertension, atherosclerosis, heart failure, myocardial infarction,
 CC ischaemic heart disease, Crohn's disease, Grave's disease, Kawasaki
 CC syndrome, Raynaud's disease, aneurysm, cerebral ischaemia, peripheral
 CC neuropathy, Alzheimer's disease, Parkinson's disease, anorexia nervosa,
 CC cachexia or diabetes. This sequence encodes the novel human glycoprotease
 CC 28472
 XX
 SQ Sequence 1821 BP; 543 A; 365 C; 394 G; 518 T; 0 U; 1 Other;
 XX
 Alignment Scores:
 Pred. No.: 1,24e-208 Length: 1821
 Score: 2088.00 Matches: 407
 Percent Similarity: 99.03% Conservative: 3
 Best Local Similarity: 98.31% Mismatches: 4
 Query Match: 98.26% Indels: 0
 DB: 10 Gaps: 0
 US-10-649-273-2 (1-414) x ABSS7020 (1-1821)
 QY 1 MetLeuIleLeuThrIleAlaGlyValAlaPhePheProSerLysArgLysValTyr 20

```

Db      147 ATGCTAATCTGACGTAAGACGTCAGAGCTTTTAAACATCAAAAAAGAAAGTTAT 206
Qy      21 GIuPheLeuAysSerPheAsnPhetiProGlyThrLeuPheLeuHisylsyleValleu 40
Db      207 GAAATTTTAAAGAGTTTAAATTTTCACTCGAACAATAATTTCTTCAATAAAATAGATTTG 266
Qy      41 GylYleGIuThrSerCyAspAspThrAlaAlaValAspGluThrGlyAsnVal 60
Db      267 GGAATTTGAATAGTGTGATGATATACAGACGCTGCTGTGTGATGATCAAACTGGAAATGTG 326
Qy      61 LeuGIuValAlaIleHisSerGIuThrGIuValHisLeuLysThrGlyIleValPro 80
Db      327 TTGGAGAAAGCAATATCTCCAAACTGAAGTTCAATTTAAAAACAGGTGGATTTGTTCT 386
Qy      81 ProAlaAlaGIuThrLeuHisArgGluAsnIleGlnArgIleValGlnGluValAlaLeuSer 100
Db      387 CCAGCAGCTCAACACCTTCAAGAGAAATAATTTCAACGAATAGTACAAAGAGCTCTTTCT 446
Qy      101 AlaseGIuValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGIuLeu 120
Db      447 GCCAGTGAAGTCTCTCCAGATGACCTCTCAGCAATTTGCACTACATAAAACAGAGACTT 506
Qy      121 AlaleuSerLeuGIuValGIuLeuSerPheSerLeuGlnLeuValGIuGlnLeuLys 140
Db      507 GCTTTAAGCTCGAGAGTGGCTTATCATTTAGCTTACCTTACAGCTGTAGAGAGTTAAAAAG 566
Qy      141 ProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLys 160
Db      567 CCATTCATTTCCATTCATCATATGAGAGGCTCATGCACTTACTATTTAGTTGTCACCAATAA 626
Qy      161 ValGIuPheProPheLeuValLeuLeuIleSerGIuValHisCysLeuLeuAlaLeuVal 180
Db      627 GTAGAATTTCTCTTTTATGTTCTTTTGAATTTCTGGAGGTCACGTCGTGTTGGCATTAAGT 686
Qy      181 GlnGIuValSerAspPheLeuLeuLeuGIuLysSerLeuAspIleAlaProGIuLysMet 200
Db      687 CAAAGAGTTTTCAGATTTTCTGCTTCTTGAAAGCTCTTTGAGACATACACAGGTGACATG 746
Qy      201 LeuAspLysValAlaArgArgLeuSerLeuIleLysIleProGIuLysSerThrMetSer 220
Db      747 CTTCAGAGGTGGCAAGAGACTTCTTAAATAAATCCAGAGGTCTCCACATGAGT 806
Qy      221 GlyGIuLysAlaIleGluHisIleLeuAlaLysGlnGIuLysAsnArgPheHisPheAspIleLys 240
Db      807 GGTGGAAAGCCATGAGAACATTTGGCCAAACAAAGAAATGATTTGATTTGACATCAAA 866
Qy      241 ProProLeuHisHisAlaLysAsnCyAspPheSerPheThrGlyLeuGlnHisValThr 260
Db      867 CCTCCCTTGATCATGCTAAATTTGTGATTTTCTTTTACTGACTTCAACACGTTACT 926
Qy      261 AspLysIleIleMetLysLysGluArgGIuGluGIuIleGluLysGIuGlnIleLeuSer 280
Db      927 GATTAATAATATGAAAAACAGAAACAAAGAGATTTGAGAAAGGGGCAAAATCTGTCT 986
Qy      281 SerAlaAlaAspIleAlaAlaThrValGlnHisIleThrMetAlaCysHisIleuValLysArg 300
Db      987 TCAGCAGCAGACATGCTGCCACAGTACGACACACATGCAATGCAATCTTGTGAAAGA 1046
Qy      301 ThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuProGlnAsnAlaVal 320
Db      1047 ACACATCGGGCTATATCTGTTTGTATGACAGAGACCTGTGTTACTCAAAATAATGCAATA 1106
Qy      321 LeuValAlaSerGIuValAlaSerAsnPhetiThrIleArgArgAlaLeuGlnIleLeu 340
Db      1107 CTGGTTGATCTGTGTGTGTGCGACATTAATCTTATATCGCAGAGGTCTGGAATTTTA 1166
Qy      341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeuCysThrAspAsnGly 360
Db      1167 ACAAAACGCAACACAGTGCATTTGTTGTGTCCTCCACAGACTATGACACTGATTAATAGGC 1226
Qy      361 IleMetIleAlaThrAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380

```

```

Db      1227 ATTAGATTCAGTCGAAATGCTATTTGAAGACTACGTCGCTGGCATTTTACATGAC 1286
Qy      381 IleGIuGlyIleArgIleGIuProLysCyProLeuGIuValAspIleSerLysGIuVal 400
Db      1287 ATTAGAGGATCCGCTATGATACCAAAATGTCTCTTGAGATGACATATCAAAAAAGAGTT 1346
Qy      401 GIuGIuAlaSerIleLysValProGlnLeuLysMetGluIle 414
Db      1347 GGAAGAGCTTCATAAAGTACCAATTAATAATGAGATA 1388

RESULT 7
ADA52832
ID ADA52832 standard; cDNA; 2208 BP.
XX
AC ADA52832;
XX
DT 20-NOV-2003 (first entry)
XX
DE Human coding sequence, SEQ ID 400.
XX
KW Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;
KW Gene Therapy; human; secretory protein; membrane proteins; cancer;
KW inflammatory disease; osteoporosis; neurological disease; gene; ss.
XX
OS Homo sapiens.
XX
PN BP1293569-A2.
XX
PD 19-MAR-2003.
XX
PP 21-MAR-2002; 2002BP-00006586.
XX
PR 14-SEP-2001; 2001JP-00328381.
PR 24-JUN-2002; 2002US-0350435P.
XX
XX
XX (HELI-) HELIX RES INST.
XX PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Negai K, Irie R, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuno Y;
XX
DR WP1; 2003-395539/38.
XX
DR P-PDB; ADA54471.
XX
XX New polynucleotides encoding full-length polypeptides, e.g. secretory
XX PT and/or membrane proteins, useful for developing medicines for diseases in
XX PT which the gene is involved, or as target molecules for gene therapy.
XX
XX Claim 1; SEQ ID NO 400; 205pp; English.
XX
XX The present invention relates to novel human secretory or membrane
XX CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
XX CC ADA54071). The coding sequences are useful in the gene therapy of
XX CC diseases caused by abnormalities of the proteins, e.g. cancer,
XX CC inflammatory diseases, osteoporosis or neurological disease.
XX
SQ Sequence 2208 BP; 660 A; 454 C; 451 G; 643 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 2,15e-193 Length: 2208
Score: 1944.00 Matches: 386
Percent Similarity: 93.24% Conservative: 0
Best Local Similarity: 93.24% Mismatches: 4
Query Match: 91.48% Indels: 24
DB: Gaps: 1

US-10-649-273-2 (1-414) x ADA52832 (1-2208)
Qy      1 MetLeuIleuThrLysThrAlaGlyValAlaPhePheLysProSerLysArgLysValIyr 20
Db      344 ATGCTAATCTGACGTAAGACGTCAGAGCTTTTAAACATCAAAAAAGAAAGTTAT 403

```

QY 21 GIUphelLeuArgSerPheAsnPhenHisProGlyThrLeuPheLeuHisIleValIleu 40
 Db 404 GAATTTTAAAGAGATTAAATTTTCACTCGAAGACATAATTTCTATATAAAATGAATG 463
 QY 41 GIYIleGluThrSerCysAspAspThrAlaAlaValValAspGluThrGlyAsnVal 60
 Db 464 GGAAATTGAACCTAGTTGTATGATACAGACGCTGCTGTGTGATGAATGAATCTGAAATGTG 523
 QY 61 LeuGIYGIUAlaIleHisSerGlnThrGluValHisLeuIleThrGlyIleValIPro 80
 Db 524 TTGGAGAAAGCAATACATTCCTCAAGCTCAAGTTCAATTTAAAAACAGGTGGATTGTTCTCT 563
 QY 81 ProAlaIleGlnGlnLeuHisIleArgIleAsnIleGlnArgIleValGlnGlnIleLeuSer 100
 Db 584 CCAGACAGCTCAACAGCTTCACAGAGAAATATTCAACGAATATGACACAGAAAGCTCTTCTCT 643
 QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleIleValProGlyIleu 120
 Db 644 GCGAGTGGAGTCTCTCAAGTACCTCTCAAGCAATGCAATGCAATCAATCAATCAATCAATCAAT 703
 QY 121 AlaLeuSerLeuGlyValGlyIleuSerPheSerLeuGlnLeuValGlyIleuLeuIleVal 140
 Db 704 GCTTTAAGCTCGAGAGTGGGCTTATCATTTAGCTTACAGCTGTAGACAGATTAAAAAG 763
 QY 141 ProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeuThrAsnIleVal 160
 Db 764 CCATTCATTCCTCATTCATCATATGAGGCTCATGCACTTACTTATGTTGATGCAATCAATCAAT 823
 QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyIleHisIleValLeuIleAlaValIle 180
 Db 824 GAGAAATTTCTTTTATGTTCTTTGATTTCTGGAGGCTCATGCTGTGTTGGATTAAGTT 883
 QY 181 GlnGlyValSerAspPheLeuLeuLeuGlyIleValSerLeuAspIleAlaProGlyAspMet 200
 Db 884 CAAAGAGTTTCAGATTTTCTGCTTCTTGAAAGCTCTTTGGACATAGACCAAGGTGACATG 943
 QY 201 LeuAspIleValAlaIleArgIleuSerLeuIleIleValHisProGluCysSerThrMetSer 220
 Db 944 CTTCACAGAGTGGCAAGAGACTTCTTTAATTAACATCCAGAGTGTCTCCACATGAGT 1003
 QY 221 GlyGlyIleValAlaIleGluHisIleuAlaIleValGlnGlyAsnArgPheHisPheAspIleVal 240
 Db 1004 GGTGGAAAGCATGAGCAATTTGGCCAAACAGAAAGAAATGATTTGATTTGACATCAACAA 1063
 QY 241 ProProLeuHisHisAlaIleValAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260
 Db 1064 CTTCCCTTGATCATGATGCTAAATATGATGATTTCTTTTACTGAGACTTCAACCGTTACT 1123
 QY 261 AspIleValIleMetIleValGlyGluGlyIleGluIleValGlyGlnIleLeuSer 280
 Db 1124 GATTAATATATATATGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 1183
 QY 281 SerAlaAlaAspIleAlaIleThrValGlnHisIleThrMetAlaCysHisIleValIleValArg 300
 Db 1184 TCAGCAGACAGATGCTGCTCCACAGTACAGCACACATAGGCAATGTCATCTTGGAAGAAAG 1243
 QY 301 ThrHisArgAlaIleLeuPheCysIleValGlnArgAspLeuLeuProGlnAsnAlaValI 320
 Db 1244 ACAACATCGGGCTATTTCTGTTTGTAGACAGAGACTTGTATCTCAAAATATATATGACAGA 1303
 QY 321 LeuValAlaSerGlyValAlaAlaSerAsnPheTyrIleArgArgAlaLeuGluIleLeu 340
 Db 1304 CTGGTTGCATCTGGTGGTGTCCGACAGTAACTTCTGATTCGACAGACTCTGGAAATTTTAA 1363
 QY 341 ThrAsnAlaIleGlnCysThrLeuLeuCysProProProArgLeuCysThrAspAsnGly 360
 Db 1364 ACAACGCAACACAGGCACTTGTGTGTCTCTCCACAGACTAGTACCTGATTAATGCG 1423
 QY 361 IleMetIleAlaIleArgIleValIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380
 Db 1424 ATTAAGATTTGCA----- 1435
 QY 381 IleGluGlyIleArgTyrGluProIleCysProLeuGlyValAspIleSerIleGluVal 400

Db 1436 -----TGATGTCCTCTTGAGTAGACATATCAAAAGAAAGTT 1471
 QY 401 GIYIleGluIleSerIleIleValProGlnLeuIleMetGluIle 414
 Db 1472 GAGAAAGCTTCATTAAGATGACCAATTAATTAATGAGATTA 1513
 RESULT 8
 ADQ24627 ID ADQ24627 standard; DNA; 2890 BP.
 XX
 AC ADQ24627:
 DT 26-AUG-2004 (first entry)
 XX
 DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 7447.
 KW soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
 KM db.
 XX
 OS Homo sapiens.
 XX
 PN MO2004048938-A2.
 PD 10-JUN-2004.
 XX
 PD 26-NOV-2003; 2003MO-US038193.
 PF 26-NOV-2002; 2002US-0429739P.
 PR 26-NOV-2002; 2002US-0429739P.
 XX
 PA (PROT-) PROTEIN DESIGN LABS INC.
 XX
 PI Aziz N, Ginsburg WM, Zlotnik A;
 XX
 DR WPI; 2004-441208/41.
 XX
 PT Early detection of soft tissue sarcoma comprises determining expression
 PT of a gene in a first soft tissue sample and a normal soft tissue sample
 PT and comparing the gene expression, also useful in treating soft tissue
 PT sarcoma.
 XX
 PS Example 2; SEQ ID NO 7447; 210bp; English.
 XX
 CC The invention relates to a novel method for detecting soft tissue sarcoma
 CC which comprises obtaining a first soft tissue sample from an individual
 CC and a normal soft tissue sample from the same or different individual,
 CC determining the expression of a gene in both samples and comparing the
 CC expression of the gene in both soft tissue samples, where a higher level
 CC of protein expression in the first soft tissue sample indicates the
 CC presence of soft tissue sarcoma. The method of the invention has
 CC cytostatic applications and may be useful for detecting soft tissue
 CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
 CC acid sequences may be useful in diagnostic and screening applications.
 CC The current sequence is that of a human soft tissue sarcoma-upregulated
 CC DNA of the invention. The current sequence is not shown within the
 CC specification per se but was submitted in CD format by the inventor.
 XX
 SQ Sequence 2890 BP; 869 A; 609 C; 611 G; 789 T; 0 U; 12 Other;
 Alignment Scores:
 Pred. No.: 3,2e-193 Length: 2890
 Score: 1944.00 Matches: 386
 Percent Similarity: 93.24% Conservative: 0
 Best Local Similarity: 93.24% Mismatches: 4
 Query Match: 91.48% Indels: 24
 DB: 12 Gaps: 1
 US-10-649-273-2 (1-414) x ADQ24627 (1-2890)
 QY 1 MetLeuIleLeuThrIleThrAlaGlyValAlaPhePheIleProSerIleValIleValIle 20
 Db 1001 ATGCTAATCTTGACTAAGATGACAGAGTGTCTTTTAAACATCAACAAAGAAAGTTAT 1060

```

Qy      21  GlupheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisIleValLeu 40
Db      1061 GAATTTTAAAGAGTTTAAATTTTCATCTGGAACACATATTTCTTCATATAAATAGATATG 1120
Qy      41  GlyIleGlyThrSerCysAspAspThrAlaAlaValValAspGlyThrGlyAsnVal 60
Db      1121 GGAATTGAACACTAGTGTGATGATATACAGACAGCTGCTGTGTGATGAACTGGAAATATG 1180
Qy      61  LeuGlyGlyAlaIleHisSerGlnThrGlyValHisIleLeuPheThrGlyGlyIleValPro 80
Db      1181 TTGGAGAGAGCAATACATTTCCAAACTGTAAGTTCATTATAAACAAGTGGAGATTGTTCT 1240
Qy      81  ProAlaAlaGlnGlnLeuHisArgGlyAsnIleGlnArgIleValGlnGlyValLeuSer 100
Db      1241 CCAGAGAGCTCAACAGCTTCAAGAGAAATATTCACGAATAGTACAAAGAGCTCTTTCT 1300
Qy      101  AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleIleProGlyLeu 120
Db      1301 GCCAGTGAAGTCTCTCAAGTGAACCTCTCAGCAATGCAACTACATATAAACAGAGACTT 1360
Qy      121  AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuIleVal 140
Db      1361 GCTTTAACCTGTGGAGTGGCTTATCTATTACCTTACAGCTGGTAGAGAGTTAAAAAG 1420
Qy      141  ProPheIleProIleHisIleMetGlyAlaHisAlaLeuThrIleArgLeuThrAsnIle 160
Db      1421 CCATTCAATCCCATTCATATATGAGAGCTCATGCACTTACATTAGCTTGACCAATATA 1480
Qy      161  ValGlyPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180
Db      1481 GTAGAAATTTCTTTTAACTTTTAACTTTTGAATTTCTGAGAGTCACTGCTGTGGCATTA 1540
Qy      181  GlnGlyValSerAspPheLeuLeuGlyValSerLeuAspIleAlaProGlyAspMet 200
Db      1541 CAAGAGATTTTCAGATTTTCTGCTTCTTGAAAGTCTTTGACATACACAGGTGACATG 1600
Qy      201  LeuAspIleValAlaArgArgLeuSerLeuIleIleHisIleProGlyCysSerThrMetSer 220
Db      1601 CTGACAAAGTGGCAAGAGACTTCTTTAATTAACATCCAGAGTCTCCACCATGAGT 1660
Qy      221  GlyGlyIleValAlaIleGlyHisIleValAlaIleGlnGlyAsnArgPheHisPheAspIle 240
Db      1661 GGTGGAAAGCCATAGAACATTTGGCCAAACAAAGAAATAGATTTCAATTTGACATCAAA 1720
Qy      241  ProProLeuHisHisAlaIleValAsnCysAspPheSerPheThrGlyLeuGlnHisIleVal 260
Db      1721 CTTCCCTTGATCATGCTGCTAAAAATTTGATTTTCTTTTAACTGAGACTTCAACACGTTACT 1780
Qy      261  AspIleIleIleMetIleValGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 280
Db      1781 GATTAATAATATATGAAAAAGAAAAAGAAAGAAAGTATGAAAGGAGGCAAAATCTGCTCT 1840
Qy      281  SerAlaAlaAspIleAlaIleAlaThrValGlnHisIleThrMetAlaCysHisIleValIle 300
Db      1841 TCAGCAGCAGACATCTCTCCACAGATACAGACACAAATGCACTGCTGTGAAAGA 1900
Qy      301  ThrHisArgAlaIleLeuPheCysIleValArgAspLeuLeuProGlnAsnAlaVal 320
Db      1901 ACACATCGGGCTATCTCTGTTTGTAAAGCAGAGACTGTTTACCTCAAAATATATGACGTA 1960
Qy      321  LeuValAlaSerGlyGlyValAlaAspAsnPheThrIleArgArgAlaLeuGlyIleLeu 340
Db      1961 CTGGTTGCACTCGTGCTGTCCGCAAGTAATCTTGATTCGCGAGAGCTCGGAAATTTTA 2020
Qy      341  ThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeuCysThrAspAsnGly 360
Db      2021 ACAAAAGCAACACAGTGCATTTGTTGTCTCTCTCCAGCATATGCACTGATATAGGC 2080
Qy      361  IleMetIleAlaIleThrAsnGlyIleGlyArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380
Db      2081 ATTTATGATTGCA----- 2092
Qy      381  IleGlyGlyIleArgGlyArgIleProIleCysProLeuGlyValAspIleSerIleGlyVal 400

```

```

Db      2093 -----NGATGTCTCTTGAGTACACATATCAAAAAGAGTT 2128
Qy      401  GlyGlyAlaSerIleIleValProGlnLeuIleValMetGlyIle 414
Db      2129 GGAAGAGCTTCCATTAAGTACCAATTAATAATGAGATTA 2170

RESULT 9
ABX70950
ID  ABX70950 standard; cDNA; 1416 BP.
XX
AC  ABX70950;
XX
05-MAR-2003 (first entry)
XX
DE  Novel human cDNA sequence #175.
XX
KW  Human; gene; ss; nervous system disorder; peripheral neuropathy;
KW  Huntington's disease; amyotrophic lateral sclerosis; haemophilia;
KW  neurodegenerative disease; Parkinson's disease; Alzheimer's disease;
KW  autoimmune disease; systemic lupus erythematosus; rheumatoid arthritis;
KW  insulin-dependent diabetes mellitus; anaemia; thrombocytopenia; wound;
KW  ulcer; burn; bone disorder; osteoporosis; osteoarthritis; stroke;
KW  fibrosis; reperfusion injury; infection; allergic rhinitis; asthma;
KW  coagulation disorder; cancer; tumour; inflammatory disease; septic shock;
KW  Crohn's disease; anaphylaxis; proliferation; chemotactic;
KW  differentiation; stem cell growth factor; haematopoiesis; chemokinetic;
KW  haemostatic; antiinflammatory; expressed sequence tag; EST.
XX
OS  Homo sapiens.
XX
PN  MO200281731-A2.
XX
PD  17-OCT-2002.
XX
PF  29-JAN-2002; 2002MO-US001222.
XX
PR  30-JAN-2001; 2001US-0074528.
XX
PA  (HYSEQ) HYSEQ INC.
PA  (GOOD) GOODRICH R W.
XX
PI  Tang TY, Liu C, Zhou P, Aeundi V, Zhang J, Zhao QA, Ren F,
PI  Xue AJ, Yang Y, Wehrman T, Wang J, Wang D, Drmanac RT;
XX
DR  WPI; 2003-058563/05.
XX
PT  Novel polypeptide useful for treating neurodegenerative diseases, myeloid
PT  or lymphoid cell disorders, bone disorders, mechanical and traumatic
PT  disorders, coagulation disorders, and inflammatory diseases.
XX
PS  Claim 1; Page; 612pp; English.
XX
XX
This invention relates to the cDNA sequences encoding an isolated novel
human polypeptide. The protein encoded by the nucleic acid of the
invention is useful for treating central and peripheral nervous system
diseases (e.g. peripheral neuropathy, Huntington's disease, amyotrophic
lateral sclerosis); neurodegenerative diseases (e.g. Parkinson's disease,
Alzheimer's disease); autoimmune disease (e.g. systemic lupus
erythematosus, rheumatoid arthritis, insulin-dependent diabetes mellitus)
CC ; myeloid or lymphoid cell disorders (e.g. anaemia and thrombocytopenia)
CC ; wounds, ulcers, burns; bone disorders (e.g. osteoporosis,
CC osteoarthritis); mechanical and traumatic disorders (e.g. stroke, head
CC trauma); lung or liver fibrosis; reperfusion injury in various tissues;
CC bacterial, viral or fungal infections; allergic conditions such as
CC allergic rhinitis, asthma; coagulation disorders (e.g. haemophilia);
CC cancer and tumours; and inflammatory diseases (e.g. septic shock, Crohn's
CC disease, anaphylaxis). The protein may be used to inhibit the growth,
CC infection or function of infectious agents such as bacteria, fungi,
CC viruses, or to effect bodily characteristics, biorythms or circadian
CC cycles of rhythms. The protein may also have
CC proliferation/differentiation, stem cell growth factor, haematopoiesis
CC regulation, immune stimulating or suppressing, chemotactic/chemokinetic,

```


CC genetic diseases, such as neurological, neuropsychological or psychiatric
CC illness. The present sequence represents murine statilycoprotease-like
CC gene sequence, which is used in an example from the present invention
XX
SQ Sequence 1572 BP, 459 A, 337 C, 340 G, 429 T, 0 U, 7 Other;

SQ Sequence 1572 BP; 459 A; 337 C; 340 G; 429 T; 0 U; 7 Other;

Alignment Scores:

Pred. No.:	3.76e-152	Length:	1572
Score:	1549.00	Matches:	326
Percent Similarity:	84.50%	Conservative:	23
Best Local Similarity:	78.93%	Mismatches:	60
Query Match:	72.89%	Indels:	7
DB:	6	Gaps:	2

US-10-649-273-2 (1-414) X ABQ75508 (1-1572)

QY	1 MetLeuullEleuthThrlYsrhrlaGlyValaPhePheLysProSerLysrArlgYsValIYr	20
Db	5 AIGCTAAATGTAAAGAAAGACAGAGAGGCTATTCCCAAGCCCCCAAGAGTAAGTTAT	64
QY	21 GluPheLeuArgSerPheasnPheHisProGlyThrlLeuPheLeuHisLysIleValLeu	40
Db	65 GGATTTTTTAAGAAAGATTAGTGTATCCACAGACTCTCTTGTGCATAAATCGTCTCG	124
QY	41 GilylIleGluThrSerCyAspAspThrIalalaIalavalalAspGluThrGlyAsnVal	60
Db	125 GGAAATGTAACACAGCGTGTATGACACAGAGGCGCGTGTGTATGATAAACCTGGGAATGTG	184
QY	61 LeuGlyGluAlaIleHisSerGlnThrGluValHisLeuLysThrGlyGlyIleValPro	80
Db	185 CTGGGGGAGCACTGCACCTCCCAACTCAGGTTCACTGAAAACAGGTGGAGTTGTTCT	244
QY	81 ProAlaIleGlnGlnLeuHisArgGluAsnIleGlnArgIleValaGlnGluAlaLeuSer	100
Db	245 CCAGTAGCTCAACAACCTTACACAGAGAAATATTCAACGATATGAAAGAAACTCTTCT	304
QY	101 AlaSerGlyValSerProSerAspLeuSerAlaIalalThrThrIleLysProGlyLeu	120
Db	305 GGCTGTAGATACACCCCAAGCCAGTCTCTACGACATTCACATTCACATCAACCGGAGCTG	364
QY	121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLys	140
Db	365 GCCCTAAGTTGGGAATGGCTTATCTTTCCTTACAGCTAGTAATCAGTTTAAARS	424
QY	141 ProPheIleProIleHisIshetGluAlaHisAlaLeuThrIleArgLeuThrAsnLys	160
Db	425 CCATTCAT -CCGATTTCATCACAGAGGCTCAGCACTGACTTTCAGCTCACCAATAAA	483
QY	161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal	180
Db	484 GTAGATTN -CCTTTTTCATGTTCTTTTGATTTCTGGGCGGACCTGCTTGGKRTTAGTC	542
QY	181 GlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet	200
Db	543 CAAGGTGTTCCGATTTCCGTCTCTTGGGAAGCTTTGGACATGACACACAGCGCATG	602
QY	201 LeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlyLysSerThrMetSer	220
Db	603 CTTGACAGGTGCAGAAAGACCTTCTTTCATCAAAACATCCAGAAATTTCTACACATAGT	662
QY	221 GlyGlyLysAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLys	240
Db	663 GGTGGAAGAGCTATAGAACAGTTGGCCAAAGACGAAATAGATTCATTTTATCATCAT	722
QY	241 ProProLeuHisHisAlaLysAsnCyAspPheSerPheThrGlyLeuGlnHisValThr	260
Db	723 CCACCTATGACAGATGCTAAGAAATTCGATTTTCTTTCACGGGACTTCAACATATTACT	782
QY	261 AspLysIleIleMetLysGlyLysGlnGlnGlyIleGlnLysGlyGlnIleLeuSer	280
Db	783 GATTAAGCTAATTAACACACAGAGAAAGAAAGAGCATTTAGAGAGGGCAAAATCTGTCA	844
QY	281 SerIleAlaAspIleAlaIalThrValGlnHisThrMetAlaCysHisLeuValLysArg	300

Db	843	TCAGCTCAGACATGCTGCTCGCGTACAGCATTCACAGCTGCCACTTGGCAAAAGA	902
Qy	301	ThrtIsaRgAlaIleLeuPheCyALysGlnARgAspLeuLeuProGlnAsnAsnAlaVal	320
Db	903	ACAGATCGCGGTATTTCTGTTTGGACAGCAAAAATTTGGCTCTCCAGCTAAACGAGTA	962
Qy	321	LeuValAlaSerGlyGlyValAlaSerAsnPheTyrIleArgAlaAlaLeuGlnIleLeu	340
Db	963	TTAGTTGATATCGAGGGTGTGGCAAGTAACCTGTGACATCCGAAAGCATTTGGAATTGTC	1022
Qy	341	ThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeuCysThrAspAsnGly	360
Db	1023	GCAAAATGCACGCATGCACGGTTGTGTGTCACCT-TCAGACCTGTGACCTGCAAAATGGG	1081
Qy	361	IleMetIleAlaIleAsnGlyIleGluArgPheArgAlaGlyLeuGlyIleLeu-HisAs	380
Db	1082	CATATGATGTGATGGATGAAGATTAAGTAAGTAAGTACGTGCTGCTTGGCCTTTACATGA	1141
Qy	380	P-IleGlyGlyIleArgTyrGluProPheCysProLeuGlyVal---AspIleSerIysG	399
Db	1142	TGATATGAGACATTCGGTATTATTAACCAAAATCTCTCTTGAGTAGAGGCATTATCCGAA	1201
Qy	399	IuValGlyGluAlaSerIleLeuValProGlnLeu	410
Db	1202	AAGTTGGCAGA--AGCTTGCCCATTAATAAAAGTTA	1233
RESULT 11			
ID	ABSt76639	standard; DNA; 1526 BP.	
XX	AC	ABSt76639;	
XX	DT	11-DEC-2002 (first entry)	
XX	DE	DNA encoding novel human metalloprotease MPI fragment #1.	
XX	XX	Metalloprotease, MP-1; immune disorder; glutamate transport; cancer;	
KW	KW	motor neuron disorder; amyotrophic lateral sclerosis; ALS; diabetes;	
KW	KW	reproductive disorder; Klinefelter's syndrome; germinal cell aplasia;	
KW	KW	genital wart; metabolic disorder; premature puberty; Kallman syndrome;	
KW	KW	Cushing's syndrome; neurodegenerative disease; Alzheimer's disease;	
KW	KW	Parkinson's disease; Huntington's disease; Tourette syndrome; sepsis;	
KW	KW	liver disease; renal disease; immune disorder; rheumatoid arthritis;	
KW	KW	acquired immunodeficiency syndrome; AIDS; pulmonary disease; pneumonia;	
KW	KW	emphysema; cystic fibrosis; vascular disorder; inflammatory disorder;	
XX	XX	neurological disorder; gene; ds.	
OS	OS	Homo sapiens.	
XX	XX		
PN	PN	WO200272751-A2.	
XX	XX		
XX	PD	19-SEP-2002.	
XX	PF	05-FEB-2002; 2002MO-US003353.	
XX	XX		
PR	PR	05-FEB-2001; 2001US-0266518P.	
XX	XX	10-APR-2001; 2001US-0282814P.	
PA	PA	(BRIM) BRISTOL-MYERS SQUIBB CO.	
XX	XX		
PI	PI	Chen J, Feder J, Nelson TC, Duclos F, Krystek S;	
XX	XX		
DR	DR	MPI; 2002-723329/78.	
XX	XX	P-PSDB; ABG96487.	
PT	PT	New isolated nucleic acid encoding MP-1 protein, useful for preventing,	
PT	PT	treating, or ameliorating diseases associated with aberrant	
PT	PT	metalloprotease activity, e.g. immune, metabolic, inflammatory and	
XX	XX	neurological disorders.	
PS	PS	Disclosure; Page 462-463; 473pp; English.	
XX	XX		

CC The invention describes an isolated nucleic acid molecule (I) encoding a
 CC metalloprotease (MP-1). (I) is useful for preventing, treating, or
 CC ameliorating a medical condition, particularly an immune disorder, an
 CC aberrant glutamate transport or motor neuron disorder, such as
 CC amyotrophic lateral sclerosis (ALS), its juvenile form or an ALS-like
 CC condition. The compositions and methods are also useful for diagnosing,
 CC prognosticating, treating, ameliorating and/or treating disorders
 CC associated with MP-1 activity, e.g. diabetes, cancer, reproductive
 CC disorders (e.g. Klinefelter's syndrome, genital warts, or germinal cell
 CC aplasia), metabolic disorders (e.g. premature puberty, Kallman syndrome,
 CC or Cushing's syndrome), neurodegenerative diseases (Alzheimer's disease,
 CC Parkinson's disease, Huntington's disease or Tourette syndrome), liver
 CC and renal diseases and immune disorders (e.g. AIDS, rheumatoid arthritis
 CC or sepsis), pulmonary diseases (e.g. pneumonia, emphysema or cystic
 CC fibrosis) and vascular, inflammatory and neurological disorders (e.g.
 CC Alzheimer's disease or Parkinson's disease). This sequence represents a
 CC metalloprotease MP1 polynucleotide
 XX

Sequence 1526 BP; 484 A; 297 C; 289 G; 456 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 5,9e-135 Length: 1526
 Score: 1385.00 Matches: 267
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 65.18% Indels: 0
 DB: Gaps: 0

US-10-649-273-2 (1-414) x ABS76639 (1-1526)

QY 148 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGlnPheProPheLeuVal 167
 Db 1 ATGGAGGCTCAGCAGCTTACTTATAGTTGACCAATAAGTAAGTAATCTTTTATGTT 60
 QY 168 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 187
 Db 61 CTTTGATTTCTGGAGGTACTGCTGTTGGCATTAAGTTCAAGAACTTTTCAGATTTTCTG 120
 QY 188 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 207
 Db 121 CTTCTTGGAAAGCTTTTGGACATAGCAGGAGGATGCTTGGACAGAGGTCGCAAGAA 180
 QY 208 LeuSerLeuIleLysHisProGlyCysSerThrMetSerGlyGlyLysAlaIleGluHis 227
 Db 181 CTTCTTTAATTAACATCAGAGTGTCTCACCATGAGTGTGGGAAAGCATAGAAACAT 240
 QY 228 LeuAlaLysGlnGlyAsnArgPheHisAspLeuIleLysProProLeuHisHisAlaLys 247
 Db 241 TTGGCAAAACAGAAATAGATTTTCAATTTGACATCAAACTCCCTTGATCATGCTTAA 300
 QY 248 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 267
 Db 301 AATTGTGATTTTCTTTTACTGTGACTTCAACACGTTACTGATTAATAATTAAGAAAG 360
 QY 268 GluLysGlnGlyGlyIleGlyLysGlyGlnIleLeuSerSerAlaAlaAspIleAla 287
 Db 361 GAAAAAGAGAAAGATTTAGAGAAAGGCAATCTCTCTTCAGCAGCAACATGCTGCC 420
 QY 288 ThrValGlnHisThrMetAlaCysHisLeuValLysValThrHisArgAlaIleLeuPhe 307
 Db 421 ACAGTACAGCACACATGTCATGTCATCTTGAAAGAAACATGCGGCTATTCTGTT 480
 QY 308 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyVal 327
 Db 481 TGTAAAGAGAGAGCTTTTACTTCAAAATTAATGAGTACATGTCATCTGTCATCTGTC 540
 QY 328 AlaSerAsnPheThrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThr 347
 Db 541 GCAAGTAACTTCTAATTCGCGAGAGCTCTGGAAATTTTAAACAAACCAACAGTGCACCT 600
 QY 348 LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaThrAsnGly 367
 Db 601 TTGTTGTCTCTCTCCAGAGTATAGCACTGATTAATGCAATTATGATTCATGAGATGCT 660

QY 368 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlnGlyIleArgTyrGlu 387
 Db 661 ATTAAGAGACTACGTCGCTGGCTGGCATTTTACATGACATAGAAAGGATCCGCTATGAA 720
 QY 388 ProLysCysProLeuGlyValAlaAspIleSerIleGlyValGlyAlaSerIleLysVal 407
 Db 721 CCAAAATGTCCTCTTGGAGTACATATCAAAAGAGTTGAGAGAGCTTCATTAAGTA 780
 QY 408 ProGlnLeuLysMetGluIle 414
 Db 781 CCACATTAATAATGAGATTA 801

RESULT 12
 ADE31345/C
 ID ADE31345 standard; DNA; 3358 BP.
 XX
 AC ADE31345;
 XX
 DT 29-JAN-2004 (first entry)
 XX
 DE Human diagnostic and therapeutic polynucleotide (dthp), SEQ ID No 100.
 XX
 KW diagnostic and therapeutic polynucleotide; dthp; antiarteriosclerotic;
 KW antiinflammatory; cerebroprotective; antihypertensive; antidiabetic;
 KW immunosuppressive; neuroprotective; nootropic; neuroleptic; tranquilizer;
 KW osteopathic; antiarthritic; antirheumatic; cyostatic; hepatotropic;
 KW virologic; haemostatic; anti-HIV; antithyroid; thyromimetic;
 KW dermatological; antibacterial; fungicide; antiparasitic; anticonvulsant;
 KW thrombolytic; anticoagulant; anorectic; vasotrophic; antileuc;
 KW gene therapy; protein replacement therapy; human; gene; ds.
 XX
 OS Homo sapiens.
 XX
 PN NO2003062376-A2.
 XX
 PD 31-JUL-2003.
 XX
 PF 13-JAN-2003; 2003WO-US001096.
 XX
 PR 16-JAN-2002; 2002US-0349384P.
 XX
 PR 17-JAN-2002; 2002US-0349413P.
 XX
 PR 17-JAN-2002; 2002US-0349946P.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 XX
 PI Jones AL, Dahl CR, Gietzen D, Chinn J, Dufour GE, Jackson JT,
 PI Yu YJ, Tuason O, Yap PE, Amshey SR, Dam TC, Liu TF, Gerstein EH;
 PI Peralta CH, Lewis SA, Chen A, Marwaha R, Lan RY, Urashka ME;
 PI Kristnam SR, Kolluru V, Panesar IS;
 DR WPI; 2003-636732/60.
 DR P-PDB; ADE31156.
 XX
 PT New human diagnostic and therapeutic polynucleotides and polypeptides,
 PT useful for diagnosing, treating or preventing e.g. leukemia, brain
 PT cancer, atherosclerosis, AIDS, thyroiditis, infections, obesity, stroke
 PT or Alzheimer's.
 XX
 PS Claim 1; SEQ ID NO 100; 634bp; English.
 XX
 CC The invention relates to a novel isolated human diagnostic and
 CC therapeutic polynucleotide (designated dthp). The novel dthp
 CC polynucleotide comprises: any of 188 DNA sequences consisting of 195-7798
 CC base pairs fully defined in the specification; a polynucleotide
 CC comprising a naturally occurring polynucleotide sequence at least 90%
 CC identical to the dthp polynucleotide; a polynucleotide complementary to
 CC the dthp polynucleotide or its polynucleotide which is at least 90%
 CC identical; or an RNA equivalent of any of the polynucleotides mentioned
 CC above. The dthp polynucleotides have the following activities:
 CC antiarteriosclerotic, antiinflammatory, cerebroprotective, antihypertensive,
 CC antidiabetic, immunosuppressive, neuroprotective, nootropic, neuroleptic,
 CC tranquilizer, osteopathic, antiarthritic, antirheumatic, cyostatic,

CC hepatotropic, vinnicide, haemostatic, anti-HIV, antithyroid, thymomimetic,
 CC dermatological, antibacterial, fungicide, antiparasitic, anticomvulsant,
 CC thrombolytic, anticoagulant, anorectic, vasotropic, and anticler. The
 CC novel DITHP polynucleotides polypeptide can be used in gene therapy and
 CC protein replacement therapy. The dithp polynucleotides or DITHP
 CC polypeptides are useful for diagnosing, preventing or treating diseases
 CC associated with the expression of human molecules. In particular, these
 CC diseases include cancers (e.g. adenocarcinoma, leukemia, melanoma, brain
 CC cancer, breast cancer, cervix cancer, bone cancer, liver cancer, lung
 CC cancer) or other cell proliferative disorders (e.g. arteriosclerosis,
 CC atherosclerosis, cirrhosis, hepatitis, polycythemia vera, primary
 CC thrombocytopenia), autoimmune/inflammatory disorders (e.g. AIDS,
 CC Addison's disease, thyroiditis, Crohn's disease, Graves' disease,
 CC Hashimoto's thyroiditis, multiple sclerosis, osteoarthritis, rheumatoid
 CC arthritis, scleroderma, systemic lupus erythematosus), infections (e.g.
 CC viral, bacterial, fungal or parasitic infection), developmental disorders
 CC (e.g. Cushing's syndrome or epilepsy), endocrine disorders (e.g.
 CC thrombosis, hypopituitarism, hypogonadism, gigantism, goiter) metabolic
 CC disorders (e.g. hypercholesterolaemia, hypoglycaemia, diabetes,
 CC hyperlipidaemia, obesity), neurological disorders (e.g. ischemic
 CC cerebrovascular disease, stroke, Alzheimer's disease, Pick's disease,
 CC Huntington's disease, Parkinson's disease, Creutzfeldt-Jakob disease,
 CC anxiety, schizophrenia), gastrointestinal disorders (e.g. ulcers),
 CC transport disorders (e.g. akinesia or multidrug resistance), or
 CC connective tissue disorders (e.g. Paget's disease or rickets). This
 CC polynucleotide sequence represents one of the human dithp DNA sequences
 CC of the invention.

XX SQ Sequence 3358 BP; 1105 A; 577 C; 601 G; 1075 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 3,39e-94 Length: 3358
 Score: 1001.50 Matches: 250
 Percent Similarity: 44.82% Conservative: 1
 Best Local Similarity: 44.64% Mismatches: 7
 Query Match: 47.13% Indels: 307
 DB: 10 Gaps: 2

US-10-649-273-2 (1-414) x ADB31345 (1-3358)

QY 157 LeuTrrasLeuValGluPheProPheLeuValLeuLeuIleSergIyGlyHisCysLeu 176
 Db :|||||
 3067 GGGACCAATTAAGTGAATTTCTTTTAAAGTCTT-TTGAATTTCTGAGGCTACTGCTG 3009
 QY 177 LeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlnGlyLysSerLeuAspIleAla 196
 Db :|||||
 3008 TGGGCAATTGTTCAAGAGATTTCAGATTTTCTGCTTGGAAAGTC-TTGAACATAGCA 2950
 QY 197 ProGlyAspMet----- 200
 Db :|||||
 2949 CCAGGAGCATGCGTTGACAGGTAATTAAGTAATTTCTCCATTTCTTTTGGTAATGT 2890
 QY 200 ----- 200
 Db 2889 TGTTCATTCACTAAGTAGCAATAGATGTCTACCACTTCACTTAATATTCTTGAA 2830
 QY 200 ----- 200
 Db 2829 TTTTATCTTAGTAACAGAACAAATTTCAATATGTTGAGAAAAATAGAAAGAGTA 2770
 QY 200 ----- 200
 Db 2769 GTACAGCAATTTTAATTTCTTACCTTTCTTAATAAATGTAAGAGTTTCATATCTGTA 2710
 QY 200 ----- 200
 Db 2709 CATTAAGCTGAATAGATTGACAGATACAGTATGTAATTTTGCCTAAATTAATGATGTA 2650
 QY 200 ----- 200
 Db 2649 AGAAGTGTCTGTAATCTAATCTACTGCAAAAAAGGTAATAAGAAATATATATAGA 2590
 QY 200 ----- 200

Db 2589 TTAACATAAGACATTAAGATGCAATGACAGAAATTAATGACACAAATTAATTACCA 2530
 QY 200 ----- 200
 Db 2529 CAGACAGGTCCCGCCGACCCCTTTGTTTAAATCTACAGAGGCTACTGCCATAT 2470
 QY 200 ----- 200
 Db 2469 ATAGAACTACAAACAACAGACAGCGTCTCCACAGTGAATTAATGAAATAGATAG 2410
 QY 200 ----- 200
 Db 2409 ACAAGTTCTTAATTAATGAGTTTCATCATTAAGCATTAATGTCACCTTCAAGCCATTT 2350
 QY 200 ----- 200
 Db 2349 CCAACCAATAGAACAGCAAAACATAGACAGGGGAGTGAATTTGCCCTTATTTGGGTC 2290
 QY 200 ----- 200
 Db 2289 ATCATAGAACAGGTTGCTGCTTACCTGAATATACGTATATATTGGCAAA 2230
 QY 200 ----- 200
 Db 2229 GTATAGCATGTTTATTTATTCATTGAGGGGTTTTTTGTTTGTAGTAATTTCAATTAATTT 2170
 QY 201 -----Leu 201
 Db 2169 CCTTTCATCTTTTCGTTTCACAGTAATTAATTTATAGCTCTAATAAATAATGTTCTTT 2110
 QY 202 AspLysValAlaArgArgLeuSerLeuIleLysHisProGluCysSerThrMetSergLy 221
 Db :|||||
 2109 GATAG-GTGGCAAGAGACTTCTTAATTAACATTCAGAGTGCTCCACATGAGTGT 2051
 QY 222 -GlyLysAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysPr 241
 Db :|||||
 2050 GGGGAAGCATAGAACATTTGSCCAACAGGAATATGATTTCAATTTTGCATCAAC 1991
 QY 241 OProLeuHisAlaLysAsnCyAspPheSerPheThrGlyLeuGlnHisValThrAs 261
 Db :|||||
 1990 TCCCTTGATCATGCTTAATAATTTGTGATTTTCTTTTACGTGACCTTCAACACGTTACTGA 1931
 QY 261 PylsIleIleMetLysLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLys 271
 Db :|||||
 1930 TAAATTAATATAGAAAAAGAAAAAGAGAGATATTTCTTAATTAAGTGAAGA 1871
 QY 272 -----GlyIleGlnLysG 276
 Db 1870 GATTAATATTCCTGATTTGCTTAATAATATAGCTGCTCATTTCTGACAGTATGAGAAAG 1811
 QY 276 LysGlnIleLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHisThrMetLacYsH 296
 Db :|||||
 1810 GGCAAATCTGTCTTCAGACAGACATTCCTGCAAGTAACAGACACATGCGATGTC 1751
 QY 296 IseLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuEupProG 316
 Db :|||||
 1750 ATCTTGGAAAGAACATCGGGCTATTCGTTTGTGAAGCAAGAGACTTGTTACTCTC 1691
 QY 316 LnaAsnAlaValLeuValAlaSerGlyValAlaSerAsnPheTyrIleArgArg 336
 Db :|||||
 1690 AAATTAATGCACTGCTGCTGATCTGT-GGTGCGCAAGTAATCTTCAATTCGCGAGAG 1632
 QY 336 LaleuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 356
 Db :|||||
 1631 CTCGGAATTTTAAC-AAAGCAACAGTGCATTTGTTGTGTCCTCCAGACTAT 1573
 QY 356 YsthrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeuG 376
 Db :|||||
 1572 GCACTGATTAATGCAATTAATGATTCATGGAATGATTTGAAAGACTAGTCTGCTGG 1513
 QY 376 LylIleuHisAspIleGlnGlyIleArgGlyArgLupProLysCysProLeuGlyValAspI 396
 Db :|||||

QY 204 ----- 204
Db 207 TTTCAATTATTCCTTGCATCTTTGCTTCACAGATATTATTTATGACTCTAAAA 148
QY 205 -----AlaArgArgLeuSerLeuIleuYshisProGluCySer 217
Db 147 AATATGTTCTTTGATAGGTGGCAGAAAGCTTTCTTATATATAACATCCAGAGTGTCC 88
QY 218 ThrMetSerGlyValYsaAlaIleGluHisLeuValAlaYsGlnGlyAsnArgPheHisPhe 237
Db 87 ACCATGAGTGTGGGAAAGCCATAGAACATTTGGCCAAACAGGAATATGATTTCAATTT 28
QY 238 AspIleLeuProProLeuHisHisAla 246
Db 27 GACATCAAACTCCCTTCATCATGCT 1
RESULT 15
ABL24632/C
ID ABL24632 standard; DNA; 3656 BP.
AC ABL24632;
XX
XX 26-MAR-2002 (first entry)
DT
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 25369.
DE
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; de.
XX
XX Drosophila melanogaster.
OS
XX WO200171042-A2.
PN
XX 27-SEP-2001.
PD
XX 23-MAR-2001; 2001WO-US009231.
PF
XX 23-MAR-2000; 2000US-0191637P.
PR 23-MAR-2000; 2000US-00614150.
PR 11-JUL-2000; 2000US-00614150.
XX
XX (PERKE) PE CORP NY.
PA
XX
PI Venter JC, Adams M, Li PMD, Myers EW;
XX
XX WPI; 2001-656860/75.
DR
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions.
XX
XX Claim 1; SEQ ID NO 25369; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins (AB57737-
CC AB572072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 3656 BP; 1035 A; 829 C; 874 G; 918 T; 0 U; 0 Other;
SQ
Alignment Scores:
Pred. No.: 3.5e-61 Length: 3656
Score: 687.50 Matches: 157
Percent Similarity: 55.64% Conservative: 60
Best Local Similarity: 40.26% Mismatches: 139
Query Match: 32.35% Indels: 34
DB: 4 Gaps: 7

US-10-649-273-2 (1-414) x ABL24632 (1-3656)
QY 39 ValLeuGlyIleGluThrSerCyAspAspThrAlaAlaValAlaAspGluThrGly 58
Db 2578 GTCTGGGCAATCGAGACTCTTGGCAGCAGCAGGAGCTGCGATGTGAGACCAAGGAGC 2519
QY 59 AsnValLeuGlyGluAlaIleHisSerGlnThrGluValHis 72
Db 2518 CGAGTATTCGCAATGCTGTGAGTCCGCAACAGAGTTCCACACCGGTGAGTGTCCACA 2459
QY 73 -----LeuYshisThr-GlyGlyIleValAlaPr 80
Db 2458 TCTCCGATGGGATGCCCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2399
QY 80 OProAlaAlaGlnGlnLeuHisAsArgGluAsnIleGluArgIleValGlnGlnIleValLeuSe 100
Db 2398 GCCCAGGAGCCAGGAGCTTCAACCCGCGCGATGAGTCCGCTTACCAAGCCGTGATGGA 2339
QY 100 rAlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleYshisProGlyLe 120
Db 2338 GCGCGCGCAATGTGAAGCCGGAACCAATGAGCGGCATGCGGCTGACACGCGTCCGAGACT 2279
QY 120 uAlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuYshis 140
Db 2278 GCGCTGAGTTGTGCTGTGGCGCTGCGCTTGGACGCGACCTGCGCGCTGCGAGAA 2219
QY 140 sProPheIleProIleHisHisMetGluAlaHisAlaLeuThrIleArgLeuThrAsn- 159
Db 2218 GCCCTGTGCTGCGCTTACCAACATGAGGCGCAGCGCGCGCGCGCGCGCGCGCGCGCG 2159
QY 160 ----LyValAlaIlePheProPheLeuValLeuLeuIleSerGlyGlyHisCyLeuLeuAl 178
Db 2158 GAGCAGATGCGCTACCCCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2039
QY 178 aLeuValGlnGlyValSerAspPheLeuLeuGlyYshisSerLeuAspIleAlaProG 198
Db 2098 GGTGGCTTAACGCGCGCGCGCGCTTACGCTTGTGGCAACATCGATGATGCGCGCG 2039
QY 198 yAspMetLeuAspIleValAlaArgArgLeuSerLeuIleYshisProGluCySerThr 218
Db 2038 CGAGGCGCTTGAACAGATCGGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTT 1979
QY 218 rMetSerGlyIleYsaAlaIleGluHisLeuAlaYshisGlnGlyValAlaAsnArgPheHis 237
Db 1978 GTGGAAGCAGAGAGGCGCGCTTACGCTTGTGGCAACATCGATGATGCGCGCG 1919
QY 237 eaSpIleYshisProProLeuHisHisAlaYshisAsnCyAspPheSerPheThrGlyLeuG 257
Db 1918 CGAGTTTCCGCTGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1859
QY 257 nHisValThrAspIleIleIleMetLeuYshisGlnGlyGlnGlyIleGluYshisGly 277
Db 1858 GAAACAATCTCTTCCG 1799
QY 277 nIleLeuSerSerAlaAlaAspIleAlaIleAlaThrValGlnHisPheMetAlaCyshis 297
Db 1798 AGCATATGAGACATACGCT 1739
QY 297 uValYshisArgThrHisArgAlaIleLeuPheCyshis-----LyGlnIleArgAspLeu 315
Db 1738 GATGACCGGAGCG 1679
QY 315 oGlnAsnAsnAlaValLeuValAlaSerGlyIleValAlaSerAsnAspThrIleArgAr 335
Db 1678 TGACACCG 1619
QY 335 gAlaLeuGluIleLeuThrAsnAlaThrGlnCyThrLeuLeuCySerProProAlaGly 355
Db 1618 CAACATGCAACATTCGCT 1559
QY 355 uCyThrAspAsnGlyIleMetIleAlaThrAsnGlyIleGluArgLeuArgAlaGlyLe 375
Db 1558 CTCTCGACACGCGCTCATGATCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1510

```

Qy      375 uGlylleuhIsaSpIleGlu---GlylleaGTyrGluProLysCyProlaenGlyVa 394
Db      1509 -----CTGCAAGATTAAGAGGCCAGCAGCGCGCTACGACTACGAC-----AGCAT 1466
Qy      394 lAspIleSerLysGluValaGlyGluAla 403
Db      1465 TGATATCCAGGGCAGCGCGGATTCGCC 1438

```

Search completed: November 10, 2005, 17:34:03
 Job time : 963.3 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: November 10, 2005, 16:53:13 ; Search time 5868 Seconds

(without alignments)
2685.516 Million cell updates/sec

Title: US-10-649-273-2

Perfect score: 2125
Sequence: 1 MLITKTAGVFFKPSKRRVY.....DISKEVGASIKVQLKMEI 414

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n.model -DEV=xlp
-O=/cgm2_1/USPFO_spool_p/US10649273/runat_02112005_091338_15562/app.query.fasta_1.1429
-DB=EST -QFMT=fastcap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10649273 @CGN 1 1 8076 @runat 02112005 091338 15562 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOC
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST: *
1: gb_est1: *
2: gb_est2: *
3: gb_hlc: *
4: gb_est3: *
5: gb_est4: *
6: gb_est5: *
7: gb_est6: *
8: gb_gest1: *
9: gb_gest2: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1827	86.0	2284	3	AK045669 Mus muscu
2	1819	85.6	1622	3	AK011265 Mus muscu
3	1634	76.9	1609	3	BC030671 Mus muscu
4	1447	68.1	870	5	BQ423651 AGENCOURT
5	1353.5	63.7	852	5	BX391919 AGENCOURT
6	1275	60.2	922	5	BQ961028 AGENCOURT
7	1173	55.2	701	2	BE740611 AGENCOURT
8	1158	54.5	749	7	CK982692 AGENCOURT
9	1144	53.8	800	7	CO738006 AGENCOURT

10	1132	53.3	1109	5	BM907988 AGENCOURT
11	1111	52.3	765	5	CF114247 ShuItzomi
12	1097	51.6	640	5	BQ636028 h03d11.Y
13	1045	49.2	730	7	CN823245 Oa.sp1bn
14	1038.5	48.9	661	7	CK833139 AGENCOURT
15	1018.5	47.9	879	5	BQ433135 AGENCOURT
16	1017	47.9	1173	6	CD508917 AGENCOURT
17	997	46.9	723	7	CK942291 AGENCOURT
18	978	46.0	579	7	CK819035 AGENCOURT
19	966	45.5	822	7	CF257246 AGENCOURT
20	963	45.3	577	7	CV028547 AGENCOURT
21	954.5	44.9	808	7	CK792978 AGENCOURT
22	954	44.9	833	5	BUI59877 AGENCOURT
23	951.5	44.8	728	4	BG436856 AGENCOURT
24	935.5	44.0	879	5	BQ256052 AGENCOURT
25	935	44.0	871	2	BF308169 AGENCOURT
26	925	43.5	658	7	CF362328 AGENCOURT
27	923	43.4	811	7	CN232675 AGENCOURT
28	920	43.3	859	5	BQ292595 AGENCOURT
29	916	43.1	831	7	CK455183 AGENCOURT
30	915	43.1	812	5	BQ246489 AGENCOURT
31	912	42.9	878	7	CK456228 AGENCOURT
32	911	42.9	661	7	CF787040 AGENCOURT
33	907	42.7	717	7	CK459683 AGENCOURT
34	906	42.6	597	6	CB272391 AGENCOURT
35	902	42.4	533	4	BM126453 AGENCOURT
36	893	42.0	574	5	BM130908 AGENCOURT
37	893	42.0	687	7	CF360919 AGENCOURT
38	879	41.4	737	7	CV073257 AGENCOURT
39	867	40.8	616	6	CB438468 AGENCOURT
40	857.5	40.4	789	5	BQ242187 AGENCOURT
41	838	39.4	637	7	CK941819 AGENCOURT
42	831	39.1	853	7	CR444993 AGENCOURT
43	825.5	38.6	806	4	BG333362 AGENCOURT
44	819.5	38.6	792	5	BQ403563 AGENCOURT
45	819	38.5	668	5	BQ285782 AGENCOURT

ALIGNMENTS

RESULT 1	AK045669	2284 bp	mus musculus adult male corpora quadrigemina cDNA, RIKEN full-length enriched library, clone: B230219017 product: similar to PUTATIVE STALOGLYCOPROTEASE TYPE 2 (Homo sapiens), full insert sequence.
ACCESSION	AK045669	GI:26337528	
VERSION	AK045669.1	GI:26337528	
KEYWORDS	HTC; CAP trapper.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	Carlinici, P. and Hayashizaki, Y.		
TITLE	High-efficiency full-length cDNA cloning		
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)		
MEDLINE	99279253		
PUBMED	10349636		
REFERENCE	2		
AUTHORS	Carlinici, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.		
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes		
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)		
MEDLINE	20499374		
PUBMED	11042159		
REFERENCE	3		
AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carlinici, P., Konno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M., Sumi, N., Iehi, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,		

Fujisawa, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
 Yonea, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuno, S., Kawai, J.,
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system -84-format
 sequencing pipeline with 384 multicapillary sequencer
 Genome Res. 10 (11), 1757-1771 (2000)
 JOURNAL
 MEDLINE
 PUBMED
 20530913
 11076861

.....
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the
TITLE PANTOM Consortium.
JOURNAL Functional annotation of a full-length mouse cDNA collection
REFERENCE Nature 405, 685-690 (2001)
AUTHORS 5
The PANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.

TITLE	Analysis of the mouse transcriptome based on functional annotation
JOURNAL	of 60,770 full-length cDNAs
REFERENCE	Nature 420, 563-573 (2002)
AUTHORS	6 (bases 1 to 2284) Adachi, J., Aizawa, K., Akimura, T., Bono, H., Carninci, P.

Fukuda, S., Furuno, M., Hanagasaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Horii, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murate, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Oso, H., Saeki, D., Shibata, K., Shingawa, A., Shiiki, T., Sano, H., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toyota, I., Yasunishi, A., Yamatsugu, M., and Hayashizaki, Y.

TITLE Direct Submission
JOURNAL Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration and Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama Kanagawa 230-0045, Japan (e-mail: genome-res@sc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

COMMENT

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site for further details.

URL:<http://genome.gsc.riken.jp/>
URL:<http://fantom.gsc.riken.jp/>

FEATURES	Location/Qualifiers
source	1. .2284

CDS

```

organism="Mus musculus"
mol_type="rRNA"
/strain="C57BL/6J"
/db_xref="PANTOM DB:B330219017"
/db_xref="taxon:10090"
/clone="B230219017"
/sex="male"
/issue_type="corpora quadrigemina"
/closure.lib="Riken full-length enriched mouse cDNA library"
/dev_stage="adult"
221..1465
/note="unnamed protein product; putative
similar to putative STATOXCOPROTEASE TYPE 2 (Homo
sapiens) (SPTRIQ9HAB0, evidence: FASTY, 80%ID, 100%length,
match=1242)"
/codon_start=1
/protein_id="BAC32450.1"
/db_xref="GI:26337529"
/translation="MMLRSTGAIKPKPSKYVGLRFSVHPTLSCHKLVLGIPTF
SCDPLGAVDEGTGALSGALSHGKGVIVPAQGLRENTQIVLEITSLGSL
RITPDSLSAINTTIKRLGALSICVGSFSLOINQPKKPIPIHMHAAHLLITLTKK
VSEGLVITIGSHCHLIVQGVSDFLKSKSLDIAGBMDLVKARSLSLIKHRECSY
ASGPAVILQIADGNKGFHTINPMQNAKCDSPGLQILITDKLITKEKEBSIENK

```

QIILSSADINAAVOHATPCHARTHTAAILFCQKILNLSPANAVLVVGGVASMLYIR
 KALEIYVNAQCTLLCPRLCTDNGIMINMNGIERLRAGLVLDVEDIIEYBKCPL
 GVDISREVAEAIKVPRLKAL"
 2262. .2267
 polyA_signal
 /note="putative"
 2284
 polyA_site
 /note="putative"
 ORIGIN

Alignment Scores:	
Pred. No.:	5.12e-194
Score:	1827.00
Percent Similarity:	90.58%
Best Local Similarity:	85.02%
Query Match:	85.98%
DB:	3
	Gaps: 0
	Length: 2284
	Matches: 352
	Conservative: 23
	Mismatches: 39
	Indels: 0
	Gaps: 0

US-10-649-273-2 (1-414) X AK045669 (1-2284)

[illegible]

Db		1001	CATTAAGCTAATTACACACAGGAAAATAAGAAGCAATTGAAAGGGCAAATCCTGTCA	1068
Oy		261	SerAlaaIAspRIleAlaAlaThrValGlnHisThrMetAlaCysHisLeuValLysArg	300
Db		1061	TCAGCTGCAGACATTGCTGCTGCCGTACAGCATGCACAGCGTCCACTTGCCAAAAAGA	1120
Oy		301	ThrHisARGAlaIleLeuPheCysLysGlnArgAspLeuLeuProGlnAsnAsnAlaVal	320
Db		1121	ACAACATCGCGCTATTCCTGTTTGGACAGAAAATTTGGTCTCTCCAGCTAACGCACATA	1180
Oy		321	LeuValAlaSerGlyGlyValAlaSerAsnPheTrpIleArgArgAlaLeuGluIleLeu	340
Db		1181	TTAGTGTGATCTGGAGGGTGGTGGCAAGTAGCTTGATCATCCGAAAAGCATTTGAATTC	1240
Oy		341	ThrAsnAlaThrGlnCysThrLeuLeuCysProProPolarGluCysThrAspAsnGly	360
Db		1241	GCAAAATGACAGCGACGTGACCGTGTGTGTGTCACCCTCCAAAGCTGTGCATGACAAATGC	1300
Oy		361	IleMetIleAlaIleAspAsnGlyIleGlnArgLeuArgAlaGlyLeuGlyIleLeuHisAsp	380
Db		1301	ATCATGATTTGATGAGTAATGAATGAATGAATGATGCTGCTGGCGGGTTCATCATAT	1360
Oy		381	IleGlnGlyIleArgTrpGlnProLysCysProLeuGlyValAspIleSerLysGlyVal	400
Db		1361	GTAGAAGACATCCCATATGAACCAAAATGCTCTTGAGATAGACATATCCAGAGAACTT	1420
Oy		401	GlyGluAlaSerIleLysValProGlnLeuLysMetGluIle	414
Db		1421	GCAGAGCTGCATPMAAGTACCGGATTAATAATGCACTT	1462
RESULT 2				
AkO11265				
LOCUS			AKO11265	1622 bp mRNA linear HTC 03-APR-2004
DEFINITION			Mus musculus 10 days embryo whole body cDNA, RIKEN full-length	
KEYWORDS			enriched library, clone:2610001m19 product:similar to PUTATIVE	
VERSION			SILOGYOOPROTEASE TYPE 2 [Homo sapiens], full insert sequence.	
ACCESSION			AKO11265	
KEYWORDS			AKO11265.1 GI:12847275	
SOURCE			HTC; CAP trapper.	
ORGANISM			Mus musculus (house mouse)	
REFERENCE			Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS			Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
TITLE			2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,	
JOURNAL MEDLINE PUBMED			High-efficiency full-length cDNA cloning	
REFERENCE			Meth. Enzymol. 303, 19-44 (1999)	
AUTHORS			99279253	
TITLE			10349636	
JOURNAL MEDLINE PUBMED			2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,	
REFERENCE			Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.	
AUTHORS			Normalization and subtraction of cap-trapper-selected cDNAs to	
TITLE			prepare full-length cDNA libraries for rapid discovery of new genes	
JOURNAL MEDLINE PUBMED			Genome Res. 10 (10), 1617-1630 (2000)	
REFERENCE			20499374	
AUTHORS			11042159	
TITLE			3 Shibata, K., Itoh, M., Mizawa, K., Nagao, S., Tasaki, N., Carninci, P.,	
JOURNAL MEDLINE PUBMED			Kono, H., Akiyama, J., Niishi, K., Kitsumaki, T., Tashiro, H., Itoh, M.,	
REFERENCE			Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,	
AUTHORS			Yamanoto, S., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,	
TITLE			Fujiwaka, R., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Wataniki, M.,	
JOURNAL MEDLINE PUBMED			Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsumura, S., Kawai, J.,	
REFERENCE			Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.	
AUTHORS			Riken integrated sequence analysis (RISA) system--384-format	
TITLE			sequencing pipeline with 384 multicapillary sequencer	
JOURNAL MEDLINE PUBMED			Genome Res. 10 (11), 1757-1771 (2000)	
REFERENCE			20530913	
AUTHORS			11076861	
TITLE			4 The RIKEN Genome Exploration Research Group Phase II Team and the	
JOURNAL MEDLINE PUBMED			PANTOM Consortium.	
REFERENCE				

	TITLE	Functional annotation of a full-length mouse cDNA collection
JOURNAL	REFERENCE	Nature 409, 685-690 (2001)
AUTHORS		5 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE		Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL		Nature 420, 563-573 (2002)
AUTHORS		6 (bases 1 to 1622) Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Fukushima, Y., Furuno, M., Hanagaki, T., Harai, A., Hayatsu, N., Hiramoto, K., Hirooka, T., Horii, F., Imotani, K., Ishii, Y., Itou, M., Izawa, M., Kakuwaka, T., Kato, H., Kawai, Y., Kojima, Y., Komio, H., Kouda, M., Koyasu, S., Kurihara, C., Matsumiya, T., Miyazaki, A., Nishik, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Salto, H., Salto, R., Sakai, C., Sakai, K., Sano, H., Saeki, D., Shibata, K., Shibata, Y., Shingawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
TITLE		Direct Submission
JOURNAL		Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-1-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-research@riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
COMMENT		Please visit our web site (http://genome.gsc.riken.jp/) for further details. cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAAGAAAGATCCAGACCTCTTTTCTTTTTTTTTCN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAAGATTCGAGTTAATTAAATTAATCCCCCCCCC 3']. cDNA was cleaved with XhoI and SacI. Cloning sites, 5' end: XhoI; 3' end: SacI. Host: SOLR.
FEATURES	SOURCE	Location/Qualifiers 1..1622 /organism="Mus musculus" /mol_type="mRNA" /cds_start="57BL/6J" /db_xref="FANTOM_DB:261000IM19" /clone="261000IM19" /tissue_type="whole body" /clone_lib="RIKEN full-length enriched mouse cDNA library" /dev_stage="10 days embryo" 207..1451 /note="unnamed protein product; putative similar to PUTATIVE SIMALOGYPROTEASE TYPE 2 (Homo sapiens) [SPTR Q9HA80, evidence: FASTV, 80%ID, 100%length, match=1242]" /codon_start=1 /protein_id="BAB27506.1" /db_xref="GI:12847276" /translation="MLMRLRTAGAIIPKPKSKYVGFLRFSEVHPRTLSCGLVLGITFTSCDDGAADVDENGVNLGSLHSOTGVLTGGIVPVPAOCLRENTORLVEITLSLSCRTIPPSLSLAATITIRGIALSLGVGSFSYLQNVQKPRTPIPHNRBAHLTRLTKRVETPLVLISGGHCILAVQGVSDPLFLKSLSLDIAEDMLDVARRLSLIKRPECTMSGSAKI EQLAKDGNRRPFTINPPMOANRCDPSFTGLQIHITDLTIHKKEGAEIEEK QILSSKDI LAAVOAHATACHLARTRHALIFCRQKNLSPANNLVSGVASNLVYR KALEIVANAOTCTLCPPPTICTDNGIMLMWGNIERLPAGLGTLHDVEDIRVPRCKPL GVDSISEVSEAALKVRRLKVAL"
polya_signal		1605..1610 /note="putative"
polya_site		1622

ORIGIN /note="putative"

Alignment Scores:
 Pred. No.: 2,38e-193 Length: 1622
 Score: 1819.00 Matches: 351
 Percent Similarity: 90.34% Conservative: 23
 Best Local Similarity: 84.78% Mismatches: 40
 Query Match: 85.60% Indels: 0
 DB: 3 Gaps: 0

US-10-649-273-2 (1-414) x AK011265 (1-1622)

QY 1 MetLeuIleuThrIleThraAlaIyAlPhePheIysProSerIysArgIysValIYr 20
 Db ATGCTAAAGTTAAGAGACAGACAGAGAGATTTCCCAAGCCCCCAAGGTAAGTTAT 266
 QY 21 GluPheLeuArgSerPheAsnPhenIspProGlyThrLeuPheLeuHisIlyValIleu 40
 Db GGATTTTAAAGAGATTAGTCTTCACTCCCAAACTCTCTCTTGTATTAACGTGCTCG 326
 QY 41 GlyIleGlyThrSerCysAspAspThrAlaAlaAlaValIAspGluThrGlyIysVal 60
 Db GGAATTTGAACACAGCTGTGATGACACAGAGCGGCTGTGTGATGAATGAACCTGGAAATG 386
 QY 61 LeuGlyIleuAlaIleHisSerGlnThrGluValHisIleuIysThrGlyIleValIPro 80
 Db CTGGGGGAAGCACTCCACTCCCAACTCAGTTCACTGAAAAACAGTGGGATGTTCTCT 446
 QY 81 ProAlaIleGlnGlnLeuHisArgGluAsnIleGlnArgIleValIleGlnGlnAlaLeuSer 100
 Db CGAGTAGCTCAACAACCTTACAGAGAAATATTCAACGAATGTAGAGAAACTCTTTCT 506
 QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleIysProGlyLeu 120
 Db GCTCTAGATGACCCCAAGCATCTCTCAGCAATTGCAACTACATCAACCGGAGCTG 566
 QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyIleuIysVal 140
 Db GCCCTAAGCTTGGAGTTAGCTTATCTTACTTACCTTACCTGTAATCAATGTTAAAG 626
 QY 141 ProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeuThrAsnIys 160
 Db CCATTCACTCCGATTCATCAACATGAGGCTCAGCACTGATTAAGGCTCAACCAATTA 666
 QY 161 ValGluPheProPheLeuValIleuLeuIleSerGlyIleHisCysLeuLeuAlaLeuVal 180
 Db GTAGATTTCTTTTATTTAGTTCTTTGATTTCTGGGCTCAGCTGTTGGCATTTAGTC 746
 QY 181 GlnGlyValSerAspPheLeuLeuGlyIysSerLeuAspIleAlaProGlyIAspMet 200
 Db CAAGGATTTCCGATTTCTGCTCTCTTGGAAAGTCTTTGGACATGACACAGCGCATG 806
 QY 201 LeuAspIysValAlaArgArgLeuSerIleIleHisIspProGluCysSerThrMetSer 220
 Db CTYGCAGAGTGGCAGAAAGACATTTCTTAATCAAACTTCAGAAATGTTCTTACAAAGAG 866
 QY 221 GlyIlyIysAlaIleGlnHisIleuAlaIysGlnGlyIysAsnArgPheHisIspheAspIle 240
 Db GGTGGAAGAGCTATAGAACAGTTGGCCAAAGACGGAATATGATTTTACTATCAAT 926
 QY 241 ProProLeuHisHisAlaIysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260
 Db CCACCTATGACAGATCTTAAGATTCGATTTTCTTTTCAAGGAGATCTTCAACATATTA 986
 QY 261 AspIysIleIleMetIysIysGlyIleGlyIleGlyIleGlyIleGlyIleGlyIle 280
 Db GATTAAGCTATATACACACAGAGAAAAAGAAAGAGCATTTAGAGAGAGCAAAATCTGTCA 1046
 QY 281 SerAlaIleAspIleAlaIleThrValGlnHisThrMetAlaCysHisIleuValIysArg 300
 Db TAGCTGCAGACATTCCTGCTGGGTACAGATGCAACAGCGTCCACCTTGGGAAAGA 1106

QY 301 ThrHisArgAlaIleLeuPheCysIysGlnArgAspLeuProGlnAsnAlaVal 320
 Db ACACATCGCGCTATTTCTGTTTTCAGACAGAAAAATTTGCTCTCTCAGCTAACGACGTA 1166
 QY 321 LeuValAlaSerGlyIleValAlaSerAsnPhenIleArgArgAlaLeuGlnIleu 340
 Db TTAGTTGTATCTGAGAGTCTTGCAGTAACTTGTATACATCCGAAAGCATTTGAAATGTC 1226
 QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeuCysThrAspAsnGly 360
 Db GCMAATGCAACGACGATGACGCTTTGTTGTCTCCCTCCAAAGACTGTGACATGACATGCG 1286
 QY 361 IleMetIleAlaThrAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380
 Db ATCATGATTCAGATGGAATGGAATTTAGTAAGTTCGTGCTGGCTTGGCCGTTTACATGAT 1346
 QY 381 IleGlnGlyIleArgGlyIleGluProIysCysProLeuGlyIleAlaIspIleSerIysGluVal 400
 Db GTTAAAGACATCCGATATGAAACCAAAATGCTCTTGAAGTACACATTCAGAGAGTT 1406
 QY 401 GlyIleAlaSerIleIysValIProGlnLeuIleuMetGluIle 414
 Db GCAGAAAGCTCCATTAAGATCCGCGATTAAATATGACACTT 1448
 RESULT 3
 BC030671 1609 bp mRNA linear HTC 19-NOV-2003
 LOCUS BC030671 Mus musculus O-sialoglycoprotein endopeptidase-like 1, mRNA (CDNA
 DEFINITION Clone IMAGE:1226118), containing frame-shift errors.
 ACCESSION BC030671
 VERSION BC030671.1 GI:2140459
 KEYWORDS HTC.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 REFERENCE
 AUTHORS Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.L., Wang J., Heien F., Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullihy J.S., Bosak S.A., McEwen P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huily S.W., Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Rahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S., Krzywinski M.I., Skalska U., Smalins D.E., Scherch A., Schein J.E., Jones S.J. and Marry M.A.
 TITLE Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 MEDLINE 22388257
 PUBMED 12477932
 REFERENCE 2 (bases 1 to 1609)
 AUTHORS Strausberg R.
 TITLE Direct Submission
 JOURNAL Submitted (20-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
 REMARK NTH-MGC Project URL: http://mgc.nci.nih.gov
 COMMENT Contact: MGC help desk
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Marcello Bento Soares, Ph.D.
 cDNA Library Preparation: M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILLUM)

DNA Sequencing by: Institute for Systems Biology
http://www.systemsbio.org

contact: amadan@systemsbiology.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/BLN at: http://image.llnl.gov
Series: IRAX Plates: 66 Row: e Column: 10
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 21312463
This clone has the following problem: frame shifted.

FEATURES

source

location/Qualifiers

1..1609
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:1226118"
/issue_type="Thymus gland, mouse"
/clone_1ib="Soares_thymus_2NMT"
/lab_host="DH10B"
/note="Vector: pT73-Pac"

ORIGIN

Alignment Scores:

Pred. No.:	1..649-172	Length:	1609
Score:	1634.00	Matches:	327
Percent Similarity:	83.86%	Conservative:	21
Best Local Similarity:	78.80%	Mismatches:	40
Query Match:	76.89%	Indels:	27
DB:	3	Gaps:	2

US-10-649-273-2 (1-414) x BC030671 (1-1609)

```
QY      1 MetLeuIleLeuThrThyThraAlaGlyValPhePheLeuProSerIysArgLysValIlyr 20
Db      229 ATGCTAAAGTTTAAAGACACAGCAGAGCTATTCACAGCCCCCAAGAGTAAAGTTTAT 288
QY      21 GluPheLeuArgSerPheAsnPhelHisProGlyThrLeuPheLeuHisIleValIleu 40
Db      289 GCAATTTTAAAGAAATTTAGTGTCTTCACTCCAGAACTCTCTTGTATTAACGTGCTCG 348
QY      41 GlyIleGluThrSerCysAspAspThrAlaAlaValAlaValAspGluThrGlyAsnVal 60
Db      349 GGAATTTGAACACAGCTGTATGATACACAGAGCGCTGTGTGATGAATCTGGGAATGTG 408
QY      61 LeuGlyGluAlaIleHisSerGlnThrGlyValHisIleu-LysThrGlyGlyIleValIle 80
Db      409 CTGGGGGAAGCACTCACTCCCAAACTCAGTTCACTGAAGT-----GGGATTCCTTC 462
QY      80 oProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSe 100
Db      463 TCCAGTACTCAACAACTTCAAGAGAAATATTCACAAATAAGTAAAGAAACCTCTTTC 522
QY      100 FAlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyIle 120
Db      523 TGCCTGTGAATCACCCCAAGGAACTCTTCAGCAATTCGACATCAACCGGAGACT 582
QY      120 uAlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLysIly 140
Db      583 GGCCCTAAGCTTGGAGTTGGCTTATCTCTTACCTTACAGCTAGTAAATCAGTTTAAAAA 642
QY      140 sProPheIleProIleHisIleMerGluAlaHisAlaLeuThrIleArgLeuThrAsnIly 160
Db      643 GCCATTCATCCCGATTCACTCATGAGGCTCAGCAGACTAGTAACTAGGCTCACCAATAA 702
QY      160 sValGluPheProPheLeuValIleLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180
Db      703 AGTAAATTCCTTTTATTTAGTCTTTTATTTCTGGCGGTCACTGCTGTGGCATTAAT 762
QY      180 lGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet 200
```

```
Db      763 CCAAGGCTTCCGATTTCTGCTCTTGGGAAGCTTTTGGACATPAGCAACGCGACAT 822
QY      200 tLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGluCysSerThrMetSe 220
Db      823 GCTTGACAGAGTGGCAAGAAAGACTTTTATTAATCAACATCCAGATTTTACATAG 882
QY      220 rGlyGlyLysAlaIleGlnHisIleuAlaLysGlnIlyAsnArgPheHisPheAspIleLy 240
Db      883 TGGTGAAAAGCTATGAAACAGTTGGCCCAAGAGCAAGAAATGATTCATTTACTATCAA 942
QY      240 sProProLeuHisIleAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValTh 260
Db      943 TCCACTGATGAGAAATGCTAAGAAATTCGATTTTCTTCACTGAGACTTCAACATATTAC 1002
QY      260 rAspLysIleIleMetLysLysGlyLysGlnGluGlyIlyIleGluLysGlyGlnIleLeuSe 280
Db      1003 TGAATACCTTAATACACACAGAGAAAGAAAGAAAGCAATTCAGAAAGGGCAATTCCTTC 1062
QY      280 rSerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCysHisLeuValLysAr 300
Db      1063 ATCAGCTGACAGACATTTCTGCTGGGTACAGATGACAGCAACAGCGTCCACTTGGCAAAAG 1122
QY      300 gThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuProGlnHisAsnAlaVala 320
Db      1123 AACACATCGCGCTAATTCGTTTTCAGAGCAAGAAATTTGCTCTTCAGCTTAACGAGAT 1182
QY      320 lLeuValAlaSerGlyValValAlaSerAsnPhetyrIleArgArgAlaLeuGluIlele 340
Db      1183 ATTAGTTGTATCTGAGAGCTTTGCAACTTCTGTACATCCGAAAGCAATTTGGAAATTTG 1242
QY      340 uThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeuCysThrAspAsnG 360
Db      1243 CGCAAAATGCAACCGAGTGCAGTGTGTTGTCTCCACCTCCAAAGCTGTGCACTGCAATGG 1302
QY      360 yIleMetIleAlaIleTrpAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAs 380
Db      1303 CATCATGATTTGCA-----TGATATCTCTTGGAGTACATATTCAGAGAAAT 1315
QY      380 pIleGluGlyIleArgTyrgLupProLysCysProLeuGlyValAspIleSerLysGlyVa 400
Db      1316 -----TGATATCTCTTGGAGTACATATTCAGAGAAAT 1350
QY      400 lGlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 414
Db      1351 TGCAGAAAGCTGCCATTAATAATACCGCATTAATAATGCACTT 1393
RESULT 4
BQ423651 870 bp mRNA linear EST 23-MAY-2002
LOCUS AGENCOURT_7790948 NIH_MGC_72 Homo sapiens cdna clone IMAGE:6065828
DEFINITION 5', mRNA sequence.
ACCESSION BQ423651
VERSION BQ423651.1 GI:21118966
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 870)
NIH-MGC http://mgs.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC/DC/DTP
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
plate: LLAM13342 row: 1 column: 21
High quality sequence stop: 710.
```


Db 797 ATTCAAGATAGTACAAAGAGCTCTTCTGCCAGTGNAGTCTCTCCAAAGTACCTCTCA 738
Qy 111 AAlaIealathrThrllelySProgllyeulaleuSereuglyVAlglyleuSerphe 130
Db 737 GCATTTGGACATACCAAGAGTCTTTTAAAGCTGGAGTGGCTTATCATTT 678
Qy 131 SerleuglnleuValglVlnleuLylySProphelleProIleHlshlsmetGluAla 150
Db 677 AGCTTACAGCTGTGAGGACAGTTAAAAAGCATTTCATTCATCATATGAGGCT 618
Qy 151 HlslaleuThrlleargleuThrasnlyValglupheProphleuValleuIle 170
Db 617 CATCACTTACTATTAGTTGACCAATTAAGTAATTTCTTTTATGTTCTTTGATT 558
Qy 171 SergllyghlySyleuLeuAlaleuValglnglyValSerAspPheleuLeuugly 190
Db 557 TCTGAGAGTCACTGCTGTTGGCATTAGTCAAGAGTTTCAGATTTCTGTTCTTGA 498
Qy 191 LysSerleuAspIleAlaProgllyAspMetleuAspLyValAlaArgArgleuSerleu 210
Db 497 AAGCTTTGACATAGCACACAGTGCATGCTTGAACAAGTGGCAAGAACTTTCTTTA 438
Qy 211 IlelyshlSProgluCySerThreSergllygllyValleIleGlnHlshleuAllys 230
Db 437 ATAAACATCCAGAGTCTCCACCATGAGTGGGAAAGCCATAGAACATTTGGCCAA 378
Qy 231 GlnGlyAsnArgPheHlSPhAspIlelySProProleuHlshlAlAllysAnCyAsp 250
Db 377 CAAGAAATAGATTTCATTTGACATCAACCTCCCTTGACATGCTTAAATTTGTAT 318
Qy 251 PheSerPheThrglyleuGlnHlshlValThrAspLyIleIleMetLylySgluLySglu 270
Db 317 TTTTCTTTACTGACCTCAACACGTTACTGATAAATATATATGAAAAAGAAAAAG 258
Qy 271 GluGlyIleGlylySgllyGlnIleleuSerSerAlaAlaAspIleAlaIleThrValGln 290
Db 257 GAAGGATTTGAGAAAGGGCAATCCTGCTTCAGACAGACATTTGCCACAGTACAG 198
Qy 291 HlshlMetAlaCyShlSleuVallySArgThrlshlSArgAlaIleleuPheCylySglu 310
Db 197 CACCAATGGCATGTCATCTTGTGAAAAGAACACATCGGCTATTCGTTTGTAAAG 138
Qy 311 ArgAspLeuLeuProgllyAsnAsnAlaValleuValAlaSergllyValAlaSerAsn 330
Db 137 AGAGACTTGTAACTCAAAATATAGCACTGCTGTCATCGTGCGTGGCCAAAGTAA 78
Qy 331 PheTyrlleArgAlaAlaGlnIleleuThrAsnAlaThrglnCyThrleuLeuCyS 350
Db 77 TTTCTATTCGCGACAGCTCTGAAATTTTAAACAAACGACACAGTGCATTTGTGTGT 18
Qy 351 ProProAspArgleu 355
Db 17 CCTCTCCACAGACTA 3

RESULT 6
LOCUS B0961028 922 bp mRNA linear EST 21-AUG-2002
DEFINITION AGENCOURT_8863711 NIH_MGC_18 Homo sapiens cDNA clone IMAGE:6423902
5', mRNA sequence.
ACCESSION B0961028
VERSION B0961028.1 GI:22376506
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 922)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: DCTD/DPF/Gazdar

CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
http://image.llnl.gov
Plate: LNCM2603 row: d column: 15
High quality sequence stop: 584.
Location/Qualifiers
1. .922
source
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6423902"
/issue_type="large cell carcinoma"
/lab_host="DH10B (phage-resistant)"
/clone_id="NIH_MGC_18"
/note="Organ: Lung; Vector: pOTB7; Site: 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."

ORIGIN
Alignment Scores:
Pred. No.: 6.75e-133 Length: 922
Score: 1279.00 Matches: 275
Percent Similarity: 91.80% Conservative: 5
Best Local Similarity: 90.16% Mismatches: 16
Query Match: 60.19% Indels: 10
DB: 5 Gaps: 3

US-10-649-273-2 (1-414) x B0961028 (1-922)

Qy 80 ProProAlAlaGlnGlnleuHlshlSArgGluAsnIleGlnArgIleValGlnGlnAlaLeu 99
Db 3 CCTCCAGAGCTTCAACAGCTTCCACAGAAATATTCAGAAATAGTACAGAGCTCTT 62
Qy 100 SerAlaSerGlyValSerProSerAspLeuSerAlleAlaThrllelySProglly 119
Db 63 TCGCCAGTGGAGTCTCTCAAGTGAACCTTCAGCAATTCACATACCAACCAAGCA 122
Qy 120 LeuAlaLeuSerleuGlyValglyleuSerPheSerleuGlnleuValgllyGlnleuLyS 139
Db 123 CTGCTTTAAAGCTGGAGTGGCTTATCATTTAGCTTACAGCTGGTAGGACAGTTAAAA 182
Qy 140 LysProPheIleProIleHlshlsmetGluAlaIleuThrlleargleuThrasn 159
Db 183 AAGCCATTCATTCATCATCATATGAGGCTCATGCACTTCTATTTAGTTGACCAAT 242
Qy 160 LysValGluPheProPheleuValleuIleSerGlyghlySyleuLeuAlaLeu 179
Db 243 AAAGTGAATTTCTTTTATGTTCTTTGATTTCTGGAGTCACTGCTGTTGGCATTTA 302
Qy 180 ValGlnGlyValSerAspPheleuLeuuglyLySerleuAspIleAlaProgllyAsp 199
Db 303 GTTCAAGAGAGTTTCAGATTTTCTGCTTTGAAAGCTTTTGAACATAGCACAGGTGAC 362
Qy 200 MetleuAspLyValAlaArgArgleuSerleuIlelyshlSProgluCyS-SerThre 219
Db 363 ATGCTTGACAGAGTGGCAAGAGACTTCTTTAATAAATCATCCAGAGTGCCTCCACAT 422
Qy 219 tSergllyLyValAlaIleGlnHlshleuAllysGlnGlyAsnArgPheHlshlSAspIle 239
Db 423 GAGTGTGGAAAGCCATGACATTTTGGCCAAACAGAAATATGATTTTCATTTTGAAT 482
Qy 239 elySProProleuHlshlAlAllysAnCyAspPheSerPheThrglyleuGlnHlshlS 259
Db 483 CAAACTCCTTGCATCATGCTTAAATTTGTGATTTTCTTTTACTGAGCTTCAACACGT 542

TITLE G.P., Bosak, S., Rubenfield, M. and Gasbarre, L.C.
Production of EST from cDNA libraries derived from immunologically
activated bovine gut
JOURNAL Unpublished (2004)
COMMENT Contact: Tad S. Sonstegard
Bovine Functional Genomics Laboratory
Animal and Natural Resources Institute
Bldg. 200 Rm2A BARC-BaBc, Beltsville, MD 20705, USA
Tel: 3015048416
Fax: 3015048414
Email: tad@anri.barc.usda.gov

Single pass sequencing. Bases called and trimmed with phred
0.000925 using options -trim_alt "-trim.fasta. Vector identified
by cross_match using options -mismatch 12 -minscore 18
Plate: 45 row: M column: 02
Seq primer: CCCAGTCACGACGCTGTGTAACG
High quality sequence stop: 749.
Location/Qualifiers
1..749
/organism="Bos taurus"
/mol_type="mRNA"
/strain="Holstein"
/db_xref="taxon:9913"
/clone="9BOV45_M02"
/sex="Male"
/tissue_type="Pooled"
/dev_stage="Multiple"
/lab_host="DH10B T1 phage resistant"
/clone_lib="BARC 9BOV"
/note="Organ: Abomasum; Vector: pagen-1; Site_1: EcorV;
Site_2: NotI; Equimolar amounts of mRNA extracted from
fundic and pyloric abomasums of 18 and 21 week old steers.
Exposure to Osteragis osteragi was initiated at 15 weeks
of age. fundic and pyloric abomasum"

ORIGIN

Alignment Scores:
Pred. No.: 2,09e-119 Length: 749
Score: 1158.00 Matches: 225
Percent Similarity: 95.18% Conservative: 12
Best Local Similarity: 90.36% Mismatches: 10
Query Match: 54.49% Indels: 0
DB: 7 Gaps: 0

US-10-649-273-2 (1-414) x CK982692 (1-749)

Qy 49 ThrTAAAlaAlaValaValaAspGluThrgLysenValLeuGluAlaIleHisSerGln 68
Db 1 ACAGCAGCTGCTGTAGTGATGAACCTGGAATGTTTGGGAGAACATATCATTCCTCAA 60
Qy 69 ThrGluValHisLeuLysThrgLysGlyIleValProProAlaAlaGlnGlnLeuHisArg 88
Db 61 ACTGAGCTCATTTTAAACAGGTGGAGTATTCCTCCAGTACGCTCAACAGCTTCCTCA 120
Qy 89 GluAsnIleGlnArgIleValGlnGlnLysSerAlaSerGlyValSerProSerArg 108
Db 121 GAAATATTCACACGATAGTACAGAGCTCTCTCCAGCAGAGTCTCTCCAGTGA 180
Qy 109 LeuSerAlaIleAlaThrThrilLysProGlyLeuSerLeuGlyValGlyLeu 128
Db 181 CTCACAGATGTGACACCATTAAGCCAGGACTTGTAAAGCTTGGGCGTAAAGTTTA 240
Qy 129 SerPheSerLeuGlnLeuValGlyGlnLeuLysLysProPheIleProIleHisSer 148
Db 241 TCGTTTACTTACACACTGTGTAGACCAAGTTTAAAGCCCTTCATTCCTCATCACAG 300
Qy 149 GluAlaHisAlaLeuThrilLysGluThrasnLysValGluPheProPheLeuValLeu 168
Db 301 GAGGCTCATGCACTCACTAATAGTTTAAACAAATAGTGAATTCCTGTTTATGTTCT 360
Qy 169 LeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 188
Db 361 TTGATTTCTGAGAGGTCATGTCTTTTGGCATTTAGTAGAGGTTTCAGATTTCTCTT 420

Qy 189 LeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgLeu 208
Db 421 CTGGAGAGCTTTGGACATAGCACCGAGTACATGCTTGAACAAGTACGAAGAAGCTT 480
Qy 209 SerLeuIleLysHisPProGluCysSerThreSerGlyGlyValAlaIleGlnHisLeu 228
Db 481 TCCCTTAATTAACATCCAGAGTGTCTCCACATAGTGTGGGAGGCTATAGAACATTTG 540
Qy 229 AlaLysGlnLysAsnArgPheHisPheAspIleLysProProLeuHisAlaLysAsn 248
Db 541 GCCAAACAGGAAATATGATTCATTTGATTTCCAGCTCCCATGCAACGCTTAAAT 600
Qy 249 CysAspPheSerPheThrgLysLeuGlnHisValThraspLysIleIleMetLysLysGlu 268
Db 601 TGTGATTTCTTTCTTTCTTGACCTTCAACAGCTTATGATTAAGATTAAGCANNAGAA 660
Qy 269 LysGluGluGlyIleGlyLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaThr 288
Db 661 AAAGAGGAGGTATCGACAGGCGCAGGTCTCTTCACTGACAGCATTTGCTGCG 720
Qy 289 ValGlnHisThrMetAlaCysHisLeu 297
Db 721 GTGCAGCACACCGTGCCTGCCACATT 747

RESULT 9
COT38006 800 bp mRNA linear EST 29-JUL-2004
LOCUS SLH03C19e02f1 squirrel heart library 1 Sperophilus lateralis cDNA
DEFINITION
Accession
COT38006
VERSION
COT38006.1 GI:50825276
KEYWORDS
EST.
SOURCE
Sperophilus lateralis (golden-mantled ground squirrel)
ORGANISM
Sperophilus lateralis
Mammalia; Eutheria; Rodentia; Sciurognathi; Sciuridae; Sciurinae;
Sperophilus.
1 (bases 1 to 800)

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Laboratory for Environmental Gene Regulation
University of Liverpool
School of Biological Sciences, The Biosciences Building, Crown
Street, Liverpool, United Kingdom, L69 7ZB
Tel: +44(0)151-795-4510
Fax: +44(0)151-795-4431
Email: cossins@liv.ac.uk
Vector has been trimmed from this EST.
Plates: 19 row: e column: 02
Seq primer: pfic T7 (5'-AATACGACTCACTAAGG-3')
High quality sequence stop: 800.
Location/Qualifiers

FEATURES

source
1..800
/organism="Sperophilus lateralis"
/mol_type="mRNA"
/db_xref="taxon:76772"
/clone="19e02"
/sex="Male & female"
/tissue_type="Heart"
/dev_stage="Adult"
/lab_host="R.coli Electromax DH10B"
/note="Vector: pFLC; Site_1: SalI GTCCAG; Site_2: BamHI
GAATC; Normalized and subcloned cDNA library prepared
from heart of hibernating and summer animals"

ORIGIN

Alignment Scores:

Pred. No.: 8.74e-118 Length: 800
 Score: 1144.00 Matches: 226
 Percent Similarity: 97.05% Conservative: 4
 Best Local Similarity: 95.36% Mismatches: 7
 Query Match: 53.84% Indels: 0
 DB: 7 Gaps: 0

US-10-649-273-2 (1-414) x C0738006 (1-800)

QY 1 MetLeuileuThrLysThrAlaGlyValPhePheLysProSerLysArgLysValTyr 20
 Db 88 ATGCTAATATTTGAATAGAACAGCGGAGTTTAAAAACATCCAAAAGGAAGTTGAT 147
 QY 21 GluPheLeuArgSerPheAsnPhenIleProGlyThrLeuPheLeuHisLysIleValLeu 40
 Db 148 GGATTTTAAAGAGTTTAAATTTTATTCCTCGAACAATAATTTCTTCAATAATAGATTG 207
 QY 41 GlyIleGluThrSerCysAspAspThrAlaAlaValValAspGluThrGlyValAsnVal 60
 Db 208 GGTATTTGAAACAAAGCTGTGATGATACAGAGCTGGGTGTGATGAAACTGGAAATGTG 267
 QY 61 LeuGlyGluAlaIleHisSerGlnThrGluValHisLeuLysThrGlyIleValPro 80
 Db 268 TTGGAGAAAGCAATACATTTCCAAACTGAAGTTCAATTTAAAAACAGTGGATTATCT 327
 QY 81 ProAlaIleGlnGluLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100
 Db 328 CCAGTAGCTCAACAATTCATAGAGAAAACATTCACGAATGATACAAAGAGCTGTTGT 387
 QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeu 120
 Db 388 GCCAGTGGAAATCTCCAAAGTGAACCTCAGAGATTGCAACTAAAGCCAGAGACTT 447
 QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLysLys 140
 Db 448 GCTTTAAGCTTAGAGTAGCTTATCATTTAGTTTACAGCTGTGATACCAAGCTGAAAG 507
 QY 141 ProPheIleProIleHisMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLys 160
 Db 508 CCATTCATCCCATTCATCATATGAGAGGCTCAGCACTTATCATTTAGTTGACCAATAA 567
 QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180
 Db 568 GTAGAGTTTCCATTTTATGTTCTTTGATTCTGGAGGTCACGTCTCTTGGCATTAAGT 627
 QY 181 GlnGlyValSerAspPheLeuLeuLeuLysSerLeuAspIleAlaProGlyAspMet 200
 Db 628 CAAAGGATTTCAGATTTTCTGCTTCTTGAAGCTTTTGGACATTTGACACAGGTGACATG 687
 QY 201 LeuAspLysValAlaIleArgArgLeuSerLeuIleLysHisProGlyCysSerThrMetSer 220
 Db 688 CTTGACAAAGTAGCAAGAAACATTTCTTTATCAACATCCAGATGCTCCACCATAGT 747
 QY 221 GlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPheHisPhe 237
 Db 748 GGGGGGAAAGGCTATAGAACATTTGGCCAAACAAAGAAACAGATTTCAATTT 798

RESULT 10
 LOCUS BM907988 1109 bp mRNA linear EST 12-MAR-2002
 DEFINITION AGENCOURT 6707465 NIH_MGC_119 Homo sapiens cDNA clone IMAGE:5744934
 5', mRNA sequence.
 ACCESSION BM907988
 VERSION BM907988.1 GI:19358367
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 1109)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.
 Email: cgs@biml.nhl.nih.gov
 Tissue Procurement: Life Technologies, Inc.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLNL12767 row: b column: 07
 High quality sequence stop: 602.
 Location/Qualifiers
 1..1109

FEATURES

source

/organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5744934"
 /issue_type="medulla"
 /lab_host="DH108"
 /clone_11b="NIH MGC 119"
 /note="Organ: Brain; Vector: pCMV-Sport6; Site 1: NotI;
 Site 2: EcoRV (destroyed); RNA source normal medulla from
 anonymous male age 27. Library is oligo-dT primed and
 directionally cloned (EcoRV site is destroyed upon
 cloning). Average insert size 1.3 kb, insert size range
 0.9-3 kb. Library is normalized and enriched for
 full-length clones and was constructed by C. Gruber
 (Invitrogen). Research Genetics tracking code 013. Note:
 this is a NIH_MGC library."

ORIGIN

Alignment Scores:

Pred. No.: 3.28e-116 Length: 1109
 Score: 1132.00 Matches: 241
 Percent Similarity: 88.89% Conservative: 15
 Best Local Similarity: 83.68% Mismatches: 23
 Query Match: 53.27% Indels: 9
 DB: 5 Gaps: 4

US-10-649-273-2 (1-414) x BM907988 (1-1109)

QY 1 MetLeuileuThrLysThrAlaGlyValPhePheLysProSerLysArgLysValTyr 20
 Db 129 ATGCTAATCTTGAATAGACTGCAAGAGTTTAAAAACATCAAAAAGGAAGTTAT 188
 QY 21 GluPheLeuArgSerPheAsnPhenIleProGlyThrLeuPheLeuHisLysIleValLeu 40
 Db 189 GAATTTTAAAGATTTTAAATTTTCAATCTGGAACACTATTTCTTCAATAATAGTATG 248
 QY 41 GlyIleGluThrSerCysAspAspThrAlaAlaValValAspGluThrGlyValAsnVal 60
 Db 249 GGAATTTAAACTAATTTGATGATGATACAGACTCTGTGTGAGTGAAGAACTGGAATGTG 308
 QY 61 LeuGlyGluAlaIleHisSerGlnThrGluValHisLeuLysThrGlyIleValPro 80
 Db 309 TTGGAGAAAGCAATACATTTCCAAACTGAATTCATTTAAAAACAGGTGGATGTTCT 368
 QY 81 ProAlaIleGlnGluLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100
 Db 369 CCACAGCTCAACAGCTTCAAGAGAAATATTCAACAAATAGTACAAAGAGCTCTTCT 428
 QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeu 120
 Db 429 GCCAGTGAAGTCTCTCCAAAGTACCTTCAGCAATTTGCACTACCATTAACCAAGACTT 488
 QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLysLys 140
 Db 489 GCTTTAAGCTTAGAGTAGCTTATCATTTAGTTTACAGCTGTGATACCAAGTAAAAAG 548
 QY 141 ProPheIleProIleHisMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLys 160
 Db 549 CCATTCATTCATTCATCATATGAGAGGCTCATGCACTTATCATTTAGTTGACCAATAA 608

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo. 1 (bases 1 to 640)
Wistow,G., Bernstein,S.L., Wyatt,M.K., Ray,S., Behal,A., Touchman,J.W., Bouffard,G., Smith,D. and Peterson,K. Expressed sequence tag analysis of human retina for the NEIBank Project: Retinding, an abundant, novel retinal cDNA and alternative splicing of other retina-preferred gene transcripts
Mol. Vis. 8 (4), 196-204 (2002)
22103461
12107411

Contact: Wistow G
Section on Molecular Structure and Function
National Eye Institute
6/331, NIH, Bethesda, MD 20892-2740, USA
Tel: 301 402 3452
Fax: 301 496 0078
Email: graeme@helix.nih.gov
Plate: 03 row: d column: 11
Seq primer: M13RP1 reverse primer (ABI).
Location/Qualifiers

FEATURES
source

1. 640
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="hd03d11"
/tissue_type="Retina"
/dev_stage="Adult"
/lab_host="EMDH108"
/clone_lib="Human Retina cDNA (Un-normalized, unamplified): hd/he"
/note="Organ: Eye; Vector: pSPORT1; Neural retina tissue was dissected from two 80 year old donors with no observed eye disease. 100ug of total RNA was used for library construction. A directionally cloned cDNA library in the pSPORT1 vector (Life Technology (Laurel MD) essentially following Bioserve Biotechnology (Laurel MD) essentially following the protocols of the SuperScript Plasmid System full details of which are contained in the manufacturer's instruction manual (http://www.lifetechn.com/). First strand synthesis was carried out using a Not I primer-adaptor
[5'-GGACTGATCTTAGATCGGAGCGCCGCC(TT)is-3']. EST analysis was performed on the unamplified library at the NIH Intramural Sequencing Center (NISC)."

ORIGIN

Alignment Scores:
Pred. No.: 1.22e-112 Length: 640
Score: 1097.00 Matches: 209
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 51.62% Indels: 0
DB: 5 Gaps: 0

US-10-649-273-2 (1-414) x BQ636028 (1-640)

QY 206 AAGAGCGleSerleuilelyshisProgluCyseSerThmetSerGlylyAlaile 225
Db 3 AAGAGCTTTCTTTAATAAACATCCAGACTGCTCCACCTGAGTGTGGGAAAGCATATA 62
QY 226 GIUHSleuAlAlysGlnGlyAnargPheHisPheaspIlelyspProPoleuHis 245
Db 63 GAACATTGGCCCAACAGGAATATGATTTCATTTTGACATCAACCTCCCTGCATCAT 122
QY 246 AAlAsnCyAspPheSerPheThrGlyLeuGlnHisValThrAspLySlelleMet 265
Db 123 GCTAAATAATGATATTTCTTTACTGACCTCAACACCTTCTGATTAATAATATATG 182
QY 266 LysLysGluLysGluLysGlylleGluLysGlyGlnleuSerSerAlaAlaAspIle 285
Db 183 AAAAAGAAAAAGAGAGGATATTGAGAGGCGCAATCTGCTTTCAGCAGCAGACATT 242

QY 286 AAlaAlarValGlnHisThrMetAlaCyshisSleuValLysArgThnHisArgAlaile 305
Db 243 GCTGCACAGTACACACACATGCAATGCTCATCTTGTAAGAAAGAACATCGGGCTATT 302
QY 306 LeuPheCyGlyGlnArgAspLeuPuroGlnAnpAnAlaValLeuValAlaSerGly 325
Db 303 CGTATTGTAACAGAGACACTGTGTAACCTCAAAATAAGCAGTACTGTTGATCTGGT 362
QY 326 GlyValAlaSerAsnPhetYrilleArgArgAlaLeuGluilleuThrAsnAlaThrGln 345
Db 363 GGTTGTGCAAGTAATCTTATATATCCGAGAGCTCTGGAAATTTTAAACGACACAG 422
QY 346 CysThrLeuLeuCyseProProProArgLeuCystrAspAsnGlylleMetIleAtrp 365
Db 423 TGCACCTTTGTGTGTCCTCTCCACAGTATGACATGATATGATATGATGATG 482
QY 366 AanglylileGluArgleuArgAlaGlyleuGlylleuHisAspIleGlyGlylleArg 385
Db 483 AATGATATTGAAGAAGATAGTGTGCTGGCATTATTACATGACATAGAGGCAATCGC 542
QY 386 TyrGluProLysCyseProLeuGlyValAspIleSerLysGlyValGlyAlaSerIle 405
Db 543 TATGAACCAAAATGTCCTTGGAGTAGACATATCAAAAGAGTTGAGAGAGCTTCAT 602
QY 406 LysValProGlnLeuLysMetGluile 414
Db 603 AAAGTACCAATTAATAATGAGATA 629

RESULT 13
CN823245 730 bp mRNA linear EST 02-JUN-2004
LOCUS
DEFINITION
Oa splbn_04N08 M13reverse Sheep spleen\brain pSPORT1 library Ovis
arles cDNA clone Oa_splbn_04N08 5', mRNA sequence.
ACCESSION
CN823245
VERSION
CN823245.1 GI:47951314
KEYWORDS
EST.
SOURCE
Ovis arles (sheep)
ORGANISM
Ovis arles
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Caprinae; Ovis.
1 (bases 1 to 730)
Gosner,A. and Hopkins,J.
Ovine spleen\brain cDNA library
Unpublished (2004)
COMMENT
Contact: J Hopkins
Veterinary Biomedical Sciences
University of Edinburgh
Summerhall Square, Edinburgh, EH9 1QH.
Email: j.hopkins@ed.ac.uk
Plate: 04 row: N column: 08
Seq primer: M13reverse
High quality sequence start: 6
High quality sequence stop: 550.
Location/Qualifiers

FEATURES
source

1. 730
/organism="Ovis arles"
/mol_type="mRNA"
/db_xref="taxon:9940"
/clone="Oa_splbn_04N08"
/clone_lib="Sheep spleen\brain pSPORT1 library"
/note="Vector: pSPORT1"

ORIGIN

Alignment Scores:
Pred. No.: 1.08e-106 Length: 730
Score: 1045.00 Matches: 204
Percent Similarity: 93.51% Conservative: 12
Best Local Similarity: 88.31% Mismatches: 15
Query Match: 49.18% Indels: 0
DB: 7 Gaps: 0

US-10-649-273-2 (1-414) x CN823245 (1-730)


```

Qy 96 GlnGluAlaLeuSerAlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThr 115
Db 37 CAAAGAGCTCTCTCTGCGCAGTAAGCTCTCCAAAGTAAGTCTCAGCAATTGCAACTACC 96
Qy 116 IleIysProGlyLeuAlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuVal 135
Db 97 GTAAGCCAGGACTTGCTTAAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAG 156
Qy 136 GlyGlnLeuLysLysProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIle 155
Db 157 GACCAAGTTAAAGACCTTCAATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 216
Qy 156 ArgLeuThrAsnLysValGlnPheProPheLeuValLeuLeuIleSerGlyGlyHisCys 175
Db 217 AGGTTAAACAATTAAGTAAGTAATTCATTTAGTCTTATTTAGTCTTATTCAGAGCTAT 276
Qy 176 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 195
Db 277 CTTTGGCATTAGCTAGAGAGAGTTTCAGATTTCTTCTTGGAAAGCTTTGGACATA 336
Qy 196 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 215
Db 337 GCACCAAGCGACATCTTGCACAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTA 396
Qy 216 CysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 235
Db 397 TGCTCCACCATGAGTGGCGGAGGAGCTATAGAACTTTGGCCAAACAGGAAATGATTTG 456
Qy 236 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerThrGly 255
Db 457 CATTTTGAATTTCCAACTCCCAACGCTGCTAAATAATTTGATTTTCTTTCTCGA 516
Qy 256 LeuGlnHisValThrAspLysIleIleMetLysLysGlnLysGlnGlyIleGluLys 275
Db 517 CTTCAACATGTTATTTGATTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTA 576
Qy 276 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleAlaThrValGlnHisThrMetAlaCys 295
Db 577 GGGAGATTCCTGCTCTTCACTGCTGACACATCTGCTGCTGCTGCTGCTGCTGCTGCTG 636
Qy 296 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 315
Db 637 CACCTTGCAAAAGAACACATCGCGCTATTTCTGCTGCAACAGAGAGCTTTGATCAT 696
Qy 316 GlnAsnAsnAlaValLeuValAlaSerGlyGly 326
Db 697 CAAAGTAACGCACTGCTGTATCTGAGGC 729

RESULT 14
CK833139 661 bp mRNA linear EST 04-MAR-2004
LOCUS DEFINITION 4057170 BARC 8BOV Bos taurus CDNA clone 8BOV_20020 5', mRNA
SEQUENCE CK833139
ACCESSION CK833139.1 GI:45061881
VERSION EST.
KEYWORDS Bos taurus (cow)
SOURCE Bos taurus
ORGANISM Bos taurus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
TITL Bovine; Bos.
1 (bases 1 to 661)
Baumann,R.G., Baldwin,R.L., Sonstegard,T.S., Van Tassel,C.P. and
Metukumalli,L.K.
Construction and Analysis of a cDNA Library Generated From
Intestinal Muscle and Epithelial Tissues of Holstein Cattle
Unpublished (2004)
CONTACT Contact: Richard G. Baumann
Bovine Functional Genomics Lab
ARRI
BLDG 162; BARC-EAST, Beltsville, MD 20705, USA
TEL: 3015048604

```

Fax: 3015048744
 Email: rbaumann@anri.barc.usda.gov
 Single pass sequencing. Bases called and trimmed with phred
 0.000925 using options -trim alt - -trim fasta. Vector identified
 by cross match using options -mismatch 12 -minscore 18
 Plate: 20 row: O column: 20
 Seq primer: CTTATTTAGGTGACATATATGAC
 High quality sequence stop: 661.
 Location/Qualifiers
 1..661
 /organism="Bos taurus"
 /mol_type="mRNA"
 /strain="Holstein"
 /db_xref="taxon:9913"
 /clone="8BOV 20020"
 /sex="Female"
 /tissue_type="Epithelial, Muscle"
 /dev_stage="Lactating, Neonatal"
 /lab_host="DH10B TONA"
 /clone_1lb="BARC 8BOV"
 /note="Organ: Intestine; Vector: pCMVSPORT6.1; Site 1:
 NCBI; Site 2: EORI; Normalized cow CDNA intestinal mRNA
 library in pCMVSPORT6.1, constructed from equimolar mRNA
 pools derived from 5 sources, 4 lactating intestinal, 1
 neonatal intestinal 4/5 Lactating, Proximal Duodenum,
 Jejunum, Distal Ileum, Colon, 1/5 Neonatal, Proximal
 Duodenum, Jejunum, Distal Ileum"

ORIGIN

Alignment Scores:

Pred. No.:	4..98e-106	Length:	661
Score:	1038.50	Matches:	203
Percent Similarity:	95.48%	Conservative:	8
Best Local Similarity:	91.86%	Mismatches:	9
Query Match:	48.87%	Indels:	1
DB:	7	Gaps:	1

US-10-649-273-2 (1-414) x CK833139 (1-661)

```

Qy 30 ProGlyThrLeuPheLeuHisLysIleValLeuGlyIleGluThrSerCysAspAspThr 49
Db 2 CCCGGG---ATATTTCTTCACTTAACATGATTTGGAAATTAACCACTTGTGATGATACA 58
Qy 50 AlaAlaAlaValAlaAspGluThrGlyAsnValLeuGlyGlnAlaIleHisSerGlnThr 69
Db 59 GAGCTGCTGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTA 118
Qy 70 GluValHisLeuLysThrGlyGlyIleValProProAlaAlaGlnGlnLeuHisArgGlu 89
Db 119 GAAGTTCATTTAAAGAACAGGTGGATTATTCCTCCAGTACTCAACGCTTCAATAGAA 178
Qy 90 AsnIleGlnArgIleValGlnGluValAlaLeuSerAlaSerGlyValSerProSerAspLeu 109
Db 179 AATATTCACGCACTATGACAGAAAGCTCTCTGCGCAGTAAGTCTCTCCAAAGTGAATC 238
Qy 110 SerAlaIleAlaThrThrIleLysProGlyLeuAlaLeuSerLeuGlyValGlyLeuSer 129
Db 239 TCAGACAGTTCACCAACCAATTAAGCCAGACTTGAAGCTTGAAGCTTGAAGCTTGAAG 298
Qy 130 PheSerLeuGlnLeuValGlyGlnLeuLysLysProPheIleProIleHisIleMetGlu 149
Db 299 TTTAGCTTCAACTGTGTAGACCAAGTTTAAAGCCCTTCACTTCCATTCATCAATGAG 358
Qy 150 AlaHisAlaLeuThrThrIleArgLeuThrAsnLysValGlnPheProPheLeuValLeu 169
Db 359 GCTCAGACACTCACTATGATTAACAATTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTA 418
Qy 170 IleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeu 189
Db 419 ATTTTGAAGGTCATTTGCTTTTGGCATTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAG 478
Qy 190 GlySerSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArgLeuSer 209

```

Db 479 GGGAAAGCTTTGACATGACACAGTGACATGCTTGAACAAGTAGCAAGAACTTCC 538
 Qy 210 LeuileuysHSPROGLUCySerThreSerGlyGlyValAlaIleGlnHisLeuAla 229
 Db 539 TTAATATAAACATCCAGTGTCTCCACCATGATGCGGGAGGCTATATGAAACATTTGGCC 598
 Qy 230 LygGlnGlyAsnAcGPhenHisPheAspIleLysProProLeuHisHisAlaLysAsnCyS 249
 Db 599 AAACAGGGAATATGATTCATTTGATTTCCAGCTCCCATGCAACGCTTAATAATTGT 658
 Qy 250 Asp 250
 Db 659 GAT 661
 RESULT 15
 BQ433135 879 bp mRNA linear EST 24-MAY-2002
 LOCUS BQ433135
 DEFINITION AGENCOURT 7760756 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:6016098
 5', mRNA sequence.
 ACCESSION BQ433135
 VERSION BQ433135.1 GI:21172211
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 879)
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cga@bbs-remail.nih.gov
 Tissue Procurement: ATCC
 CNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LIML at:
 http://image.llnl.gov
 Plate: LLM13213 row: d column: 19
 High quality sequence stop: 674.
 FEATURES
 source location/Qualifiers
 1..879
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone_image:6016098"
 /issue_type="embryonal carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_92"
 /note="Organ: testis; Vector: PCWV-SPORT6; Site: 1: NotI;
 Site 2: SalI; Cloned unidirectionally; oligo-dT primed.
 Average insert size 2.5 kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH_MGC Library."
 ORIGIN
 Alignment Scores: 1.39e-103 length: 879
 Pred. No.: 1018.50 Matches: 210
 Score: 94.67% Conservative: 3
 Percent Similarity: 93.33% Mismatches: 9
 Best Local Similarity: 47.93% Indels: 3
 Query Match: 5 Gaps: 1
 DB: 5
 US-10-649-273-2 (1-414) x BQ433135 (1-879)
 Qy 1 MetLeuileuThryStrThralagIValPhePheLysProSerLysArgLysValTyr 20
 Db 201 ATGCTATATCTTGACTAAGACGTGAGGAGTTTATTAACCATCAAAAAAGAAAGTTTAT 260
 Qy 21 GluPheLysGSerPheAsnPhenHisProGlyThreLysPheLysHisLysIleValLeu 40

Db 261 GAATTTTAAGAGTTTAAATTTTCACTCTGGAACATAATTTCTTCAATAAATAGTATG 320
 Qy 41 GIIYIEGLuThSerCyAspAspPhrAlaAlaValAlaValAspGluThrGlyAsnVal 60
 Db 321 GGAATTTGAACATGATTTGATATATACAGACGCTGCTGTGTGATGAATGAACATGAATG 380
 Qy 61 LeuGlyGlyAlaIleHisSerGlnThrGlyValHisLeuLysThrGlyGlyIleValPro 80
 Db 381 TTGGGGAAGCAATACATTTCCCAACTGAAGTTCAATTTAAAAACAGGTGGGATTTGCTCT 440
 Qy 81 ProAlaAlaGlnGlnLeuHisArgGlyLysIleGlnArgIleValGlnGlnAlaLeuSer 100
 Db 441 CCAGCAGCTCAACAGCTTCACAGAGAAATATTCACGATAGTACAGAAAGCTCTTTCT 500
 Qy 101 AlaserGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeu 120
 Db 501 GCCAGTGGAGTCTCTCCAGATGACCTCTCAGCAATTCGCAATCAATCAATCAATCAAT 560
 Qy 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLysLys 140
 Db 561 GCTTTAAGCTGGGAGTGGGCTTATCATTTAGCTTACAGCTGGTGAAGACAGTTAAAAAG 620
 Qy 141 ProPheIleProIleHisHisMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLys 160
 Db 621 CCATTCATTTCCCATCATATATGAGAGGCTCATGCACTTATATAGTTGACCAATATAA 680
 Qy 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180
 Db 681 GTAGAAATTTCTTTTATGTTCTTTGATTTCTGAGAGTCACTGTCGTTGGCATTTGCT 740
 Qy 181 GlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet 200
 Db 741 CAAGAGTTTCAGATTTTTCGCTCTTGGAAAGTCTTTGACATAGCACAGGATGACATG 800
 Qy 201 -LeuAspLysValAlaArgArgLeuSerLeuIle---LysHis-ProGluCysSerThrm 219
 Db 801 GCTTGACAAGGTGGGCAAGAAAGACTTTCTTAAATATAAATTCCTCAAGTGCCTCCCC 860
 Qy 219 etSerGlyGly 222
 Db 861 ATGAANGGT 871

Search completed: November 11, 2005, 02:14:03
 Job time : 5885 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 10, 2005, 16:54:39 ; Search time 293.85 Seconds

(without alignments)
2305.321 Million cell updates/sec

Title: US-10-649-273-2

Perfect score: 2125

Sequence: 1

BLAST62
Xgapop 10.0 , Xgapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-O=/cgn2_1/USPTO_spool_p/US10649273/rnat/02112005_091339_15581/app_query.fasta_1.1429
-DB=Issued_Patents_NA -QWRT=fastcap -SUFFIX=p2n.rn1 -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pct -NORF=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10649273 @CGN 1.1.177 @rnat/02112005_091339_15581 -NCPU=6 -ICPU=3
-NMAP -LARGESQERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOC
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:*

1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2125	100.0	2197	4	US-10-067-443-1
2	2090.5	98.4	1387	4	US-10-067-443-21
3	1747	82.2	1416	4	US-09-774-528-177
4	1385	65.2	1526	4	US-10-067-443-23
5	1186.5	55.8	14364	4	US-10-067-443-20
6	502	23.6	94750	4	US-09-596-002-38
7	501	23.6	1053	4	US-09-540-236-806
8	499.5	23.5	1206	4	US-09-252-991A-801
9	492.5	23.2	1059	4	US-09-252-991A-884
10	482.5	22.7	1074	4	US-09-543-681A-2341
11	482	22.7	996	4	US-09-902-540-6612
12	482	22.7	2582	4	US-09-902-540-503

13	479	22.5	1315	1	US-08-087-797-1	Sequence 1, Appli
14	475.5	22.4	1830121	4	US-09-557-884-1	Sequence 1, Appli
15	475.5	22.4	1830121	4	US-09-643-990A-1	Sequence 1, Appli
16	474.5	22.3	1032	4	US-09-489-039A-2050	Sequence 2050, Ap
17	472	22.2	1092	4	US-09-107-532A-2955	Sequence 2955, Ap
18	461	21.7	1008	3	US-08-987-121A-5	Sequence 5, Appli
19	461	21.7	1011	3	US-08-987-121A-3	Sequence 3, Appli
20	457	21.5	1006	3	US-08-961-083-51	Sequence 51, Appli
21	457	21.5	1006	3	US-09-536-784-51	Sequence 51, Appli
22	457	21.5	10974	3	US-08-961-527-214	Sequence 214, App
23	455	21.4	1011	3	US-09-066-512-1	Sequence 1, Appli
24	454	21.4	1011	3	US-09-134-000C-1551	Sequence 1551, Ap
25	453	21.3	1011	4	US-09-583-110-2196	Sequence 2196, Ap
26	448	21.1	1011	4	US-09-107-433-1618	Sequence 1618, Ap
27	443	20.8	1107	3	US-09-134-001C-1072	Sequence 1072, Ap
28	433	20.4	1026	3	US-09-149-624-1	Sequence 1, Appli
29	424	20.0	15249	4	US-08-966-717E-102	Sequence 102, App
30	424	20.0	15249	4	US-08-781-986A-102	Sequence 102, App
31	406.5	19.1	640681	4	US-09-790-988-1	Sequence 1, Appli
32	406.5	19.1	1230025	4	US-09-438-185A-1	Sequence 1, Appli
33	406.5	19.1	1230025	4	US-09-602-777A-147	Sequence 147, Appli
34	404.5	19.0	1155	4	US-09-438-185A-1	Sequence 1, Appli
35	404	19.0	3064	3	US-09-221-017B-794	Sequence 794, App
36	379.5	17.9	4403765	3	US-09-103-840A-2	Sequence 2, Appli
37	379.5	17.9	4411529	3	US-09-103-840A-1	Sequence 1, Appli
38	369.5	17.4	645	4	US-09-328-352-483	Sequence 483, App
39	362	17.0	876	4	US-09-724-623-19	Sequence 19, Appli
40	357	16.8	580073	4	US-08-545-528D-1	Sequence 1, Appli
41	333.5	15.7	42325	4	US-08-311-731A-131	Sequence 131, App
42	327.5	15.4	3993	4	US-09-710-279-3985	Sequence 3985, Ap
43	319	15.0	822	4	US-09-710-279-727	Sequence 727, App
44	317	14.9	36941	4	US-08-311-731A-130	Sequence 130, App
45	313	14.7	1664976	4	US-08-916-421B-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-10-067-443-1
; Sequence 1, Application US/10067443
; Patent No. 6642041
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN
; TITLE OF INVENTION: SPINAL CORD, MP-1
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 2197
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (231) ..(1472)
US-10-067-443-1
Alignment Scores:
Pred. No.: 5, 72e-261
Score: 2125.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 100.00%
DB: 4
Gaps: 0
US-10-649-273-2 (1-414) x US-10-067-443-1 (1-2197)

QY 1 MetLeu1leuThrLySerThrAlaGlyValPhePheLySProSerLySArgLySValTyr 20
Db 231 ATGCTAACTTCACTAAGCTCAGAGAGTTTAAAAACATCAAAAGAAAGTTAT 290
QY 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisLyS11ValLeu 40
Db 291 GAATTTTAAAGAAAGTTTAAATTTTCTCGAACACTATTTCTTCATATAAATAGATATG 350
QY 41 Gly11leG1uThrSerCySAspSPrThrAlaAlaValValAspG1uThrGlyAsnVal 60
Db 351 GGAATTGAAACTAGTGTGATGATACAGACAGCTGCTGTGTGATGATAAAGCTGAAATGTG 410
QY 61 LeuG1yGluAla11eh1sSerG1nThrGluValHisLeuLySThrG1yG11ValPro 80
Db 411 TTGGAGAAAGCAATACATTCCTCCAAAGTCAAGTTCAATTTAAAAACAGTGGAGTTCTCT 470
QY 81 ProAla1a1aG1nG1nLeuHisArgG1uAsn11eG1nArg11eValG1nG1uAlaLeuSer 100
Db 471 CCAGACAGCTCAACAGCTTCACAGAGAAATATTCAACGAATAGTACAGAGAGCTTTTCT 530
QY 101 AlaSerGlyValSerProSerAspLeuSerAla11eAlaThrThr11eLySProG1yLeu 120
Db 531 GCCAGTGAAGTCTCCCAAGTACCTCTCAGCAATTCGCAATTAACCAAGAGACTT 590
QY 121 AlaLeuSerLeuG1yValG1yLeuSerPheSerLeuG1nLeuValG1yG1nLeuLyS1yS 140
Db 591 GCTTTAAGCTCGAGAGTGGCTTTATCATTTAGCTTACAGCTGTAGAGACAGTTAAAAAG 650
QY 141 ProPhe11ePro11eh1s1sMetG1uAlaHisAlaLeuThr11eArgLeuThrAsnLyS 160
Db 651 CCATTCATTCCTCCATTCATATAGAGGCTCATGACATCTACTATAGTTAGTTGCAATAAA 710
QY 161 ValG1uPheProPheLeuValLeuLeu11eSerG1yG1yHisGyS1eLeuAlaLeuVal 180
Db 711 GTAGAAATTTCTTTTATTTAGTTCTTTTGTGATTTCTGGAGGTCACGTCTGTGGCATTA 770
QY 181 G1nG1yValSerAspPheLeuLeuG1yLyS1eLeuAsp11eAlaProG1yAspMet 200
Db 771 CAAAGAGTTTCAGATTTCTGCTTCTTGAAAGCTTTTGACATAGCACAGGTTGACATG 830
QY 201 LeuAspLySValAlaArgArgLeuSerLeu11eLySHisProG1yCyS1eThrMetSer 220
Db 831 CTTGACAGAGTGGCAAGAGACTTTCTTTATATAAATCCAGAGGCTCCACACAGAGT 890
QY 221 G1yG1yLySAla11eG1uHis1eLeuAlaLyS1eG1nG1yAsnArgPheHisPheAsp11eLyS 240
Db 891 GGTGGAGAAAGCCATATAGACATTTGGCCAAACAGAAATAGATTCAATTTTGACATCAA 950
QY 241 ProProLeuHisHisAlaLySAsnCySAspPheSerPheThrG1yLeuG1nHisValThr 260
Db 951 CTTCCCTTGCAATCTAATAAATGTGATTTTCTTTTACTGAACTTCAACAGTTACT 1010
QY 261 AspLyS11e11eMetLyS1yS1yG1uLyS1yG1uG1y11eG1uLyS1yG1n11eLeuSer 280
Db 1011 GATTAATAATATATGAAAAAGAAAAAGAAAGATTTGAGAGGGCAAAATCTCTCT 1070
QY 281 SerAlaAlaAsp11eAlaAlaThrValG1nHis1eThrMetAlaCySHis1eLeuValLySArg 300
Db 1071 TCAGCAGCAGACATCTCTCCACAGTACAGCACACATGCGCATGCTCTTGAGAAAGA 1130
QY 301 ThrHis1sArgAla11eLeuPheCyS1yS1yG1nArgAspLeuLeuProG1nAsnAlaVal 320
Db 1131 ACAACATCGGGCTAATTTCTTTTGTAAAGCAGAGAGACTTTGTACTCAAAATATAGCGTA 1190
QY 321 LeuValAla1eSerG1yG1yValAlaSerAspPheTyr11eArgArgAlaLeuG1u11eLeu 340
Db 1191 CTGGTTCATCTGT 1250
QY 341 ThrAsnAla1eThrG1nCyS1eThrLeuLeuCySProProArgArgLeuCyS1eThrAspAsnG1y 360
Db 1251 ACAAAAGCAACACAGTGCATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1310
QY 361 11eMet11eAla1eThrAsnG1y11eG1uArgLeuArgAlaG1yLeuG1y11eLeuHisAsp 380

Db 1311 ATTATGATTCAGAGATGATATGAAAGACTAGTCTGCTGGCATTTTACATGAC 1370
QY 381 11eG1yG1y11eArgArgG1uProLyS1yCySProLeuG1yValAsp11eSerLyS1yVal 400
Db 1371 ATAGAAAGCAATCCGCTATGAAACCAAAATGTCTTGTGAGTACATATCAAAAGAGTT 1430
QY 401 GlyGluAlaSer11eLySValProG1nLeuLyS1eMetG1u11e 414
Db 1431 GGAAGAGCTTCATTAAGATACCAATTTAAAAATGAGATA 1472
RESULT 2
US-10-067-443-21
; Sequence 21, Application US/10067443
; Patent No. 6642041
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE HIGHLY EXPRESSED
; TITLE OF INVENTION: SPINAL CORD, MP-1
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 21
; LENGTH: 1387
; TYPE: DNA
; ORGANISM: homo sapiens
US-10-067-443-21
Alignment Scores:
Pred. No.: 6,62e-257 Length: 1387
Score: 2090.50 Matches: 412
Percent Similarity: 93.85% Conservative: 0
Best Local Similarity: 93.85% Mismatches: 2
Query Match: 98.38% Indels: 25
DB: 4 Gaps: 1
US-10-649-273-2 (1-414) x US-10-067-443-21 (1-1387)
QY 1 MetLeu1leuThrLySerThrAlaGlyValPhePheLySProSerLySArgLySValTyr 20
Db 24 ATGCTAACTTCACTAAGCTCAGAGAGTTTAAAAACATCAAAAGAAAGTTAT 83
QY 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisLyS11ValLeu 40
Db 84 GAATTTTAAAGAAAGTTTAAATTTTCTCGAACACTATTTCTTCATATAAATAGATATG 143
QY 41 Gly11leG1uThrSerCySAspSPrThrAlaAlaValValAspG1uThrGlyAsnVal 60
Db 144 GGAATTGAAACTAGTGTGATGATACAGACAGCTGCTGTGTGATGATAAAGCTGAAATGTG 203
QY 61 LeuG1yGluAla11eh1sSerG1nThrGluValHisLeuLySThrG1yG11ValPro 80
Db 204 TTGGAGAAAGCAATACATTCCTCCAAAGTCAAGTTTAAAAACAGTGGAGATGTCTCT 263
QY 81 ProAla1a1aG1nG1nLeuHisArgG1uAsn11eG1nArg11eValG1nG1uAlaLeuSer 100
Db 264 CCAGACAGCTCAACAGCTTCACAGGAAATATTTCAACGAATAGTACAGAGAGCTTTCT 323
QY 101 AlaSerGlyValSerProSerAspLeuSerAla11eAlaThrThr11eLySProG1yLeu 120
Db 324 GCCAGTGAAGTCTCTCCAGAGTCACTCCACGAATTCGACATTAACCAATTAACAGAGACTT 383
QY 121 AlaLeuSerLeuG1yValG1yLeuSerPheSerLeuG1nLeuValG1yG1nLeuLyS1yS 140
Db 384 GCTTTAAGCTCGAGAGTGGCTTTATCATTTTGTGATTAACAGCTGTGAGGACAGTTAAAAAG 443
QY 141 ProPhe11ePro11eh1s1sMetG1uAlaHisAlaLeuThr11eArgLeuThrAsnLyS 160

Db 880 GAGAGGGGCAAAATCTGCTTTCAGCAGACAGATTCGTCACAGTACAGCACACAAATG 939
Qy 294 A|A|C|y|s|h|i|s|l|e|u|v|a|l|y|s|a|g|t|h|r|h|i|s|a|g|a|l|l|e|u|p|h|e|c|y|a|l|y|s|g|i|n|a|s|e|r|a|s|p|h|e|u 313
Db 940 GCATGCTATCTTGTGTAAGAAACACATCGGGCTATTCGTGTTGTGTAAGAGAGCTTG 999
Qy 314 LeuProGlnAsnAlaValLeuValAlaSerGlyValAlaSerAsnPhetYrile 333
Db 1000 TTACTCTAAATATATGACAGTACTGCTTCATCTGATGCTGGTGGCAAGTAACTTATATAC 1059
Qy 334 ArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProPro 353
Db 1060 CGCAGAGCTCTGGAATTTTAAACCAACGACAGTGCATTTGTGTCTCTCTCC 1119
Qy 354 ArgLeuCysThrAsnAsnGlyIleMetIleAlaThrAsnGlyIleGlnArgLeuAla 373
Db 1120 AGACTATACCTGTAATGAGCACTTATGATGATGATGAAATGATTAAGACTACCTGCT 1179
Qy 374 G|l|y|e|u|g|i|l|e|u|h|i|s|a|p|i|l|e|g|i|l|e|a|g|y|l|e|a|g|y|l|e|u|p|h|e|c|y|s|p|r|o|l|e|u|g|i|y 393
Db 1180 GCGTTGGCATTTCATGACATGACATGAGGATCCGCTATGAAACCAAAATGCTCTTGA 1239
Qy 394 ValAspIleSerLysGluValGlyGlnAlaSerIleLysValProGlnLeuLysMetGlu 413
Db 1240 GTAGCATATCAAAAGAAAGTGGAGAGCTTCATTAAGTACCAATTAATAATGAGAG 1299
Qy 414 I|l|e| 414
Db 1300 ATA 1302

RESULT 4
US-10-667-443-23
; Sequence 23, Application US/10067443
; Patent No. 6642041
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE HIGHLY EXPRESSED IN
; TITLE OF INVENTION: SPINAL CORD, MP-1
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 23
; LENGTH: 1526
; TYPE: DNA
; ORGANISM: homo sapiens
US-10-667-443-23

Alignment Scores:
Pred. No.: 1 62e-166 Length: 1526
Score: 1385.00 Matches: 267
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 65.18% Indels: 0
DB: 4 Gaps: 0

US-10-649-273-2 (1-414) x US-10-667-443-23 (1-1526)

Qy 148 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGlnPheProPheLeuVal 167
Db 1 ATGGAGGCTCATGCCATTACTATAGTTGACCAATTAAGTATGATTTCTTTTATGTT 60
Qy 168 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 187
Db 61 CTTTGATTTTCGAGAGTCACTGCTGTGGCATTTGATTCAGAGAGTTTTCAGATTTTCTG 120
Qy 188 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 207

Db 121 CTTCTGGAAAGTCTTTGACATAGACCAAGTGAATGCTTTTACAAAGTGGCAAGAGA 180
Qy 208 LeuSerLeuIleLysHisProGluCysSerThrMetSerGlyLysAlaIleGluHis 227
Db 181 CTTTCTTAATTAATAATCATCAGAGTCTCCACCATATGATGTGTGGAAAGCATATGACAT 240
Qy 228 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 247
Db 241 TTGGCCAAACCAAGAAATAGATTTCATTTGATCATCAAACTCCCTTGATCATATGCTAAA 300
Qy 248 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 267
Db 301 AATTTGATTTTCTTTTACTGTGACTTCACACGTTATCTATTAATTAATTAAGAAAAAG 360
Qy 268 GluLysGlnGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAla 287
Db 361 GAAAAAGAGAGATTTTGAAGAGGGCAAAATCTGCTTCAGAGAGACATTTGCTGCC 420
Qy 288 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 307
Db 421 ACAGTACAGACACAAATGCGATGTCATCTTGAAAAAGAACACATCGGGCTATTCGTGTT 480
Qy 308 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyVal 327
Db 481 TGTAAAGAGAGACATTTGTTAATCTCAAAATTAATGACAGTACTGTTGATGCTGTGTC 540
Qy 328 AlaSerAsnPhetYrileArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThr 347
Db 541 GCAAGTACTTCTATATCCGAGAGCTCTGGAATTTTAACCAAGCAGACAGTGCAT 600
Qy 348 LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaTrpAsnGly 367
Db 601 TTGTTGTGCTCTCTCCACATATGACATGATATGCAATTAATGATGCAATGAAATGCT 660
Qy 368 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyGlu 387
Db 661 ATTTGAAAGACTACGCTGCTGGCATTTTATCATACATAGAGCATCCGCTATGAA 720
Qy 388 ProLysCysProLeuGlyValAspIleSerLysGluValGlyGlnAlaSerIleLysVal 407
Db 721 CCAAAATGCTCTTGGAGTACATATCAAAAGATGAGAGCTTCATTAAGTA 780
Qy 408 ProGlnLeuLysMetGluIle 414
Db 781 CCACATTAATAAATGAGATTA 801

RESULT 5
US-10-667-443-20
; Sequence 20, Application US/10067443
; Patent No. 6642041
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE HIGHLY EXPRESSED IN
; TITLE OF INVENTION: SPINAL CORD, MP-1
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 20
; LENGTH: 14364
; TYPE: DNA
; ORGANISM: homo sapiens
US-10-667-443-20

Alignment Scores:
Pred. No.: 2,19e-139 Length: 14364
Score: 1186.50 Matches: 313
Percent Similarity: 32.30% Conservative: 1

Best Local Similarity: 32.20% Mismatches: 2
Query Match: 55.84% Indels: 657
DB: 4 Gaps: 4
US-10-649-273-2 (1-414) x US-10-067-443-20 (1-14364)

Qy 74 LysThrGlyGlyIleValProProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArg 93
Db 10623 AGAACAGGTGGATGTTCTCCACAGCTCAACACCTTCACAGAAAAATATTCACAG 10682

Qy 94 IleValGlnGluAlaLeuSerAlaSerGlyValSerProSerAspLeuSerAlaIleAla 113
Db 10683 ATAGACAAAGAGCTCTTCTCCAGTGAGTCTCCCAAGTACCTTCACAGATTGCA 10742

Qy 114 ThrThrIleLysProGlyLeuAlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGln 133
Db 10743 ACTACCATTAACCGAGCTTCTTAAGCTGGGAGTGGGCTTATCATTTAGCTTACAG 10802

Qy 134 LeuValGlyGlnLeuLysLysProPheIleProIleHisHisMetGluAlaHisAlaLeu 153
Db 10803 CTGGTAGACAGCTTAAAAAGCATTCATCCCATTCATCATATGAGGCTCATGCACTT 10862

Qy 154 ThrIleArgLeuThrAspLysValGluPheProPheLeuValLeuLeuIleSerGlyGly 173
Db 10863 ACTATAGGTGACCAATTAAGTAGAATTTCTTTTATGTTCTTTGATTTCTGGAGGT 10922

Qy 174 HisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeu 193
Db 10923 CACTGTCTGTTGGCATTTAGTTCAAGAGTTTCAGATTTCCTGCTTCATGGAAGTCTTGG 10982

Qy 194 AspIleAlaProGlyAspMetLeuAspLysVal ----- 204
Db 10983 GACATAGACACGAGTGACATGCTTGCAAGGT -AATTAGAAATTAATTTCTCATTTCTT 11041

Qy 204 ----- 204
Db 11042 TTTGTATGTTGTCATTCACTAAGTAGCAATAGATGTGTACCACCTTCACTTAA 11101

Qy 204 ----- 204
Db 11102 TATTTTCGAATTTTATCTTAGTAACTGAAAAAATTCACATATGTGTAGAAAAATAGA 11161

Qy 204 ----- 204
Db 11162 AAGAGTAGTACACATTTTATTAATTTAGCCTTTCTTAATAAATGTAGAAGTTCAAT 11221

Qy 204 ----- 204
Db 11222 ATCTGTACATAAAGCTGAAAAATAGTTTGCAATACAGTATATTTTGGCAATATATGT 11281

Qy 204 ----- 204
Db 11282 ATGTGAAGAAGCGTCTGTGAACATACTGCAGAAAAAGTAAATAAGAGATAT 11341

Qy 204 ----- 204
Db 11342 ATATAGATTACATAGGACATTAAAGATGCAATGCAAGATTAAATTCACACATTTACT 11401

Qy 204 ----- 204
Db 11402 TACACACAGACAGGGTCCCCCCCCCTTTGTTTAGATATAGTACAGAGGCTACT 11461

Qy 204 ----- 204
Db 11462 GCCATATATAGAAAAACAACAACAACAACAACAACATGCTTCCACAGTGAATTA 11521

Qy 204 ----- 204
Db 11522 TAGAGATATAGACAAAGTCTTATTTATTTAGCGTTCAATCATTAAGCAGTTATTTGCACT 11581

Qy 204 ----- 204
Db 11582 TCAGCCCATTTTCCAAACCAATAGAGAAAGCAACATAGACAGGGGCGAGTTGGCTCT 11641

Qy 204 ----- 204
Db 11642 TATTTGTCGGGTGATCATTAAGAAACAGGGTTGTCTGCTTACCTGAATATCAGTATAGTC 11701

Qy 204 ----- 204
Db 11702 TATATTTGCCAAGATATAGCATGTTTATTCATTCAGGGGTTTTTTGTTTGTATAT 11761

Qy 204 ----- 204
Db 11762 TTTCAATTTATTTCTTTGTCATCTTTTCGTTTCCAGATATTTAATTTATGACTTAA 11821

Qy 205 ----- 205
Db 11822 AATATGTTCTTTGATAGTGGGAAAGACCTTTCTTAATTAACATCCAGAGTGTCC 11881

Qy 218 ThrMetSerGlyGlyValAlaIleGluHisLeuAlaLysGlnGlyAsnArgPheHisPhe 237
Db 11882 ACCATGAGTGTGGGAAAGCCATAGAACATTTGGCCAAACAGGAAATAGATTTCAATTTT 11941

Qy 238 AspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGln 257
Db 11942 GACATCAACCTCTCCCTTGACATGCTTAAATAATTTGATTTTCTTTTACTGACCTCAA 12001

Qy 258 HisValThrAspLysIleIleMetLysLysGlyLysGlu ----- 271
Db 12002 CACTTACTGATTAATTAATTAATGAAGAAAGAAAAAGAGAGATATTTCTTAATTAAT 12061

Qy 272 ----- 272
Db 12062 AAGTTGAACAGATTAATATATCTCGATGTGCTTAAATAAATAGCTGCTCATTTCTGACGG 12121

Qy 272 YllleGlyLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHisThr 292
Db 12122 TATTTGAAGGGGCAATATCTGCTTCAGACAGACATTTGCTGCCACAGTACACGACAC 12181

Qy 292 rMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAs 312
Db 12182 AATGCAATGTCATCTTTGTAAAGAACACATCGGGCTATTTCTGTTGTAGACAGAGGA 12241

Qy 312 PleuLeuProGlnAsnAsnAlaValLeu ----- 321
Db 12242 CTTGTTACCTCAAAATATATGCACTACTGTGAAGTTTATGCTCATTTATAGTAATAGTTA 12301

Qy 321 ----- 321
Db 12302 CACTTGCATATATGTTACTTTTTCACAGACCTTGCCTTGTGTAGATGAACAGAT 12361

Qy 321 ----- 321
Db 12362 CTTTATGCTTATGTAAGCTGACAGTATGAAATTAATGACGATAGAAAGCTAACAG 12421

Qy 321 ----- 321
Db 12422 CCAATTTCTGTACTAGTTTGTAGCTTTAATGAGACAGCTGTATAGCTTTTATAGCACATA 12481

Qy 321 ----- 321
Db 12482 AGTCTAATTTGCACTCTCTGTGTGATTTTAAAGAGGCTTACAAATTAAGAAAGTAAAT 12541

Qy 321 ----- 321
Db 12542 GCAGTAAGCTATATCATATTTTAAAGAAATATAGTGATTTCTTCATCTTTGATGA 12601

Qy 321 ----- 321
Db 12602 ATCCCTTGTGTGTTGTTTATTTTAAATTAAGCCAGTCAATTTTACAGTGGAGGTGAT 12661

Qy 321 ----- 321
Db 12662 TCCAACTTTCGTGACATAATATGTATTAAGTTCTGATATATCACTATATTTGACAGCC 12721

```
Qy 321 ----- 321
Db 12722 AAAATCCCTTAATTGCTTAAAAAGCCTTGACAAACATCCGTTTAACGTATCTTAA 12781
Qy 321 ----- 321
Db 12782 CTTTATTCATTTAAAAATTATTAACCTAAAGTGGAAGAAAGTTTAAATGAGTAAATTCAT 12841
Qy 321 ----- 321
Db 12842 AGATGAAATTTTACATGATATCAAGAAATATTTTTCAGAGTTATGATGATAAATGCA 12901
Qy 321 ----- 321
Db 12902 CAAATATATTAATAATTTTACGGCTCTAAATATGATCTATGATGAAATTAATTAATA 12961
Qy 321 ----- 321
Db 12962 AATATTTAGATGAAGTTGGAAGAAATATATCAAAATATGTAATGTTGTATGCTA 13021
Qy 321 ----- 321
Db 13022 TTGAATATATTAATTAATTTTCTTCCAAATTTTATATACATAGATATGTCATCTGCC 13081
Qy 321 ----- 321
Db 13082 CATTACCATCTCAAAATGGAGTATGTTATTAATGCTGATATTTTCTCCAGG 13141
Qy 321 ----- 321
Db 13142 TTTAATTAGACGCTTGTCATATCCATATATAGTATTTGTTTCTTCAATTCCT 13201
Qy 322 ---ValAlaSerGlyIleGluThrSerAspPheTyrIleArgGalaLeuGluIleLeu 340
Db 13202 TCAGGTCATCTGGTGGTGGTGGCAAGTAACTTCTATATCCGACAGGCTCTGGAATTTT 13261
Qy 340 uThraSnaIaThrGlnCysThrLeuLeuCysProPheArgIleuGluThrAspSnaI 360
Db 13262 AACAAACCCACACAGTGCATCTTGTGTCTCTCCACAGCATATGACATGATATATG 13321
Qy 360 YIleMetIleAlaTyr----- 365
Db 13322 CATTTATGATTCATGATGTAAGCCACAGATATACGTCTTCACTCAATACATATTAATAT 13381
Qy 365 ----- 365
Db 13382 TAATTGCAATTTTATCATACTAAGCCTTCTCTTCAGATCTTGAGCTATTGATTTTAT 13441
Qy 366 -----AenGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuH 379
Db 13442 TTTAATGCTTCTTAATTAGAAATGATTAAGAAAGACTACGTGCTGGCTTGGCATTTTTAC 13501
Qy 379 IAspIleGluGlyIleArgTyrGluProIys 389
Db 13502 ATGACATAGAAAGCATTCGCTATGAACCAAG 13533

RESULT 6
US-09-596-002-38
; Sequence 38, Application US/09596002
; Patent No. 6632636
; GENERAL INFORMATION:
; APPLICANT: Lagace, Robert, E.
; APPLICANT: Paterson, Chandra
; APPLICANT: Berg, Kim, L.
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME
; FILE REFERENCE: PM-0008-4 US
; CURRENT APPLICATION NUMBER: US/09/596,002
; CURRENT FILING DATE: 2000-06-16
; PRIOR APPLICATION NUMBER: 60/140,121
; PRIOR FILING DATE: 1999-06-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PERL Program
; SEQ ID NO 38
```

```
; LENGTH: 94750
; TYPE: DNA
; ORGANISM: M. catarrhalis
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte template ID No. 6632636 38
; PUBLICATION INFORMATION:
US-09-596-002-38

Alignment Scores:
Pred. No.: 2,66-50 Length: 94750
Score: 502.00 Matches: 129
Percent Similarity: 50.81% Conservative: 59
Best Local Similarity: 34.86% Mismatches: 149
Query Match: 23,62% Indels: 34
DB: 4 Gaps: 9

US-10-649-273-2 (1-414) x US-09-596-002-38 (1-94750)
Qy 39 ValLeuGlyIleGluThrSerCysAspPheThrAlaAlaValValAspGluThr--- 57
Db 23873 GATTTGGATTTGGAGCATCTTGTGATGAACAGGCGTCCATCTATGATGATGATG 23932
Qy 58 -----GlyAsnValLeuGlyGluAlaIleHisSerGlnThrGluValHisLeu 73
Db 23933 AATGGCGTGTGGCGCGGCTGTCTCAGGTTTGTATGACCAATTAATTAATCAATGCC 23992
Qy 74 LysThrGlyGlyIleValProPheAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArg 93
Db 23993 ACCTATGGCGGTGTCTGTGCTGAGCTTGCACGTGACCACTTCGTAAAGCTTGTGCCG 24052
Qy 94 IleValGlnGluAlaLeuSerAlaSerGlyValSerProSerAspLeuSerAlaIleAla 113
Db 24053 TATTTAATGAGCTGTGTGATCAGGCAATATATCAAAATCCATATGATGCGGTGCCG 24112
Qy 114 ThrThrIleLysProGlyLeuAlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGln 133
Db 24113 TATACCAAGGCGCAGGCGCTGATTTGGGCGCTGTATGACAGGGCATTAATTTGGCGGACG 24172
Qy 134 LeuValGlyGlnLeuLysLysProPheIleProIleHisIleMetGluAlaHisAlaLeu 153
Db 24173 CTGGCGTATGGCGTGGCGCGTCCAGCGTTGGGTGATCATATGAGGAGCATCTGTTA 24232
Qy 154 ThrIleArgLeuThr-----AsnLysValGluPheProPheLeuValLeuIleSer 171
Db 24233 GCAACCGTATTTGCCAGTATGACCCCAAGCTTTCGTTGTGTGCTGCTGCTGTGTCG 24292
Qy 172 GlyIleHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLys 191
Db 24293 GGGCGTATACATCATGCTGTGCGCGCGGATGATGCGGTGCGGTATCATGATTTGGCGGAG 24352
Qy 192 SerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgGluSerLeuIle 211
Db 24353 TCTATGATGATGCGGTGGGTGATGATTAACCGCAAAATGCTCAAACTG--- 24409
Qy 212 LysHisProGluCysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGln 231
Db 24410 CCTATATCT-----GTrGGCCCAATATATGAAAAATTTAGCCAAAAAC 24451
Qy 232 GlyAsnArgPheHisPheAspIleLysProProLeuHisIleAlaLysAsnCysAspPhe 251
Db 24452 GGCACACCAACAGCGCTATGAGCTGCCAAGACCCATGACAGAT---AAAGGCTGTGATTTT 24508
Qy 252 SerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysGluLysGluGlu 271
Db 24509 TCGTTGATGGCATGAAGAACCGCATTCATATCTCAACAAAGACACCAAGCCCA 24568
Qy 272 GlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHis 291
Db 24569 AGCAGACCC-----GCCACACGACCAAGCATCGCGCAAGCTTTGAGATAT 24613
Qy 292 ThrMetAlaCysHisIleValIleArgThrHisArgAlaIleLeuPheCysLysGluArg 311
DB: 9
```


Db 24614 GCGGTGGTGGACTTTGGTCAAAAATGCAACCAACTACGATGACAGGCATTCC 24673
Qy 312 AspleuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyValAlaSerAspHe 331
Db 24674 CAG-----CTGGTGTGCGAGGGGGCGCTCTGCGCAATCAG 24709
Qy 332 TyrIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCyPro 351
Db 24710 ATGCTACCGCCGACCTCTGACCGAGACGCTCCGCCAATGCATGCGTGTACTATGCC 24769
Qy 352 ProProArgLeuCyethrAspAsnGlyIleMetIleAlaThrAsnGlyIleGluArgLeu 371
Db 24770 CCGACCGACCTATGACAGGATTAATGCGATGATGCTGCTGCTGCTTTGTCGGCTC 24829
Qy 372 ArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrGluProLysCyPro 391
Db 24830 AGCTGTGGACAGTCG-----GATGACTTGGC-GGTTCCGTGATTCC-----CCG 24873
Qy 392 LeuGlyValAspIleSerTyrGluValGly 401
Db 24874 ATGGGATATGACAGCGCTTGGCGTATCGGC 24903
RESULT 7
US-09-540-236-806
Sequence 806, Application US/09540236
Patent No. 6673910
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATAR
FILE REFERENCE: 2709.2005-001
CURRENT APPLICATION NUMBER: US/09/540,236
CURRENT FILING DATE: 2000-04-04
NUMBER OF SEQ ID NOS: 3840
SEQ ID NO 806
LENGTH: 1053
TYPE: DNA
ORGANISM: M.catarrhalis
US-09-540-236-806
Alignment Scores:
Pred. No.: 1,23e-53 Length: 1053
Score: 501.00 Matches: 126
Percent Similarity: 49.86% Conservative: 56
Best Local Similarity: 34.52% Mismatches: 139
Query Match: 23.58% Indels: 44
Gaps: 8
US-10-649-273-2 (1-414) x US-09-540-236-806 (1-1053)
Qy 39 ValLeuGlyIleGluThrSerCysAspAspThrAlaAlaValAlaSerGluThr--- 57
Db 16 GTATTGGATTGAGACATCTTGATGAACAGGGCTTGCCATCATATAGTACGATG 75
Qy 58 -----GlyAsnValLeuGluIleValAlaIleHisSerGlnThrGluValHisLeuLys 74
Db 76 AATGCGAGGGGGGGGGTGTCTGTCTACAGTTTATATAGCCAATTAATATACATGCGACC 135
Qy 75 ThrGlyGlyIleValProProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIle 94
Db 136 TATGCGGCTGTGCTGCTGACCTGCGACGACCAATTCGTAAGCTTGGCCGTTA 195
Qy 95 ValGlnGluAlaLeuSerAlaSerGlyValSerProSerAspLeuSerAlaIleAlaThr 114
Db 196 TTTAATGAGATCTTGATCAAGCAATATACCAATTCGACATTCGTCGGTGGTAT 255
Qy 115 ThrIleLysProGlyLeuAlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeu 134
Db 256 ACCAAAGGCCCGGGCTGATTGGGGCATTTGATGACAGGGGCAATTATTGGCGGACCGTG 315
Qy 135 ValGlyGlnLeuLysProPheIleProIleHisIleMetGluAlaHisAlaLeuThr 154
Db 316 GCGTATGGGCTTGGCGGTGACGCGGTGGGCGTGCATCATATGAGGAGACATCTTTAGCA 375

Qy 155 IleArgLeuThr-----AsnLysValGluPheProPheLeuValLeuLeuIleSerGly 172
Db 376 CCGTATTGGCCAGATGATGACCCCCAAGCTTTCCGTTTGTGCTGCTGCTGCTGCGGC 435
Qy 173 GlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSer 192
Db 436 GGTGATCATGCTGCTGCTGCGCGCGATGTGGCGGCTGTATGATGATATGAGTGGGAGTCC 495
Qy 193 LeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgAspLeuSerLeuLys 212
Db 496 ATGATATATGCGGTGGCGGAGTCTTTGATTAACGGCAAAATGCTTAACTG---CCC 552
Qy 213 HisProGluCysSerThrMetSerGlyLysAlaIleGlnHisLeuAlaLysGlnGly 232
Db 553 TATCCT-----GGTGTCCAAATATCGAAAATTAATGCAAAAAGCGC 594
Qy 233 AsnArgPheHisPheAspIleLysProProLeuHisHisValAlaLysAsnCyAspPheSer 252
Db 595 AACCCACACCGCTATGAGCTGCGCCAAAGACCGATGACGAT---AAAGGCTGGATTTCG 651
Qy 253 PheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLysGluGluGly 272
Db 652 TTCAGTGGCATGAAAACCGCATTCATATCTCAAAAGACACGCCAAACGCCAAAGC 711
Qy 273 IleGluLysGlyIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnIleThr 292
Db 712 GACCT-----GCCACACAGACACATCCCGGACCTTTGATATGCG 756
Qy 293 MetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAsp 312
Db 757 GTGTGGATCTTTGTTGCAAAAATGACCAAGGACATGACATGACAGGATTCGCGCAG 816
Qy 313 LeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyValAlaSerAspPheTyr 332
Db 817 -----CTGGTGTGCGAGGGGGCGCTCTGCGCAATCAGACG 852
Qy 333 IleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCyProPro 352
Db 853 CTACGGCCGACCTCTGACCGGAGACGCTCGGCCAATGCATGCGTGGTCTACGCGCCG 912
Qy 353 ProArgLeuCyThrAspAsnGlyIleMetIleAlaThrAsnGlyIleGluArgLeuArg 372
Db 913 ACCGAGTATGACAGGATATGATGATGATGATGCTTACCGCTTTGTCGGCTAACG 972
Qy 373 AlaGly-----Leu 375
Db 973 CGTGGCAGTCGATGACTTGGCGGTGCTGTATTCGCCAGATGATGACGATGCTT 1032
Qy 376 GlyIleLeuHisAsp 380
Db 1033 GGTATCGAATATGAT 1047
RESULT 8
US-09-252-991A-801
Sequence 801, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 801
LENGTH: 1206
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-801

Alignment Scores:

Pred. No.:	2,436-53	Length:	1206
Score:	499.50	Matches:	133
Percent Similarity:	51.68%	Conservative:	67
Best Local Similarity:	34.37%	Mismatches:	158
Query Match:	23.51%	Indels:	29
		Gaps:	10

US-10-649-273-2 (1-414) x US-09-252-991A-801 (1-1206)

```

Oy 13 LysProSerLysArgLysValTyrGluPheLeuArgSerPheAsnPhetHisProGlyThr 32
Db 91 AAAGCTCCGATGCAAAACACTGCTGTGGCCAAAGCCCGCATTTATCATGGCGGCTT 150
Oy 33 LeuPheLeuHisLysIle-----ValLeuGlyIleGluThrSerGys 46
Db 151 CGTATTTGGCAAGCTCTAAGGCTCAAGCCCATGCGGCTGTGGAGACTGAAAGCTCTGC 210
Oy 47 AsparThrAlaAlaValAlaValAspGluThrGlyAsnValLeuGlyGluAlaIleHis 66
Db 211 GACGAAACCGGCGTCCCTTACGACGACGAAACGCGGCTGTGCGACGCGCTGTTC 270
Oy 67 SerGluThrGluValHisLeuLysThrGlyGlyIleValProProAlaIleGluGlnLeu 86
Db 271 AGTCAGATGACCTTCATCGCGTCAAGCGGCGGTGTGTCGCCAGCTGGCTTGGCGCGAC 330
Oy 87 HisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSerAlaSerGlyValSerPro 106
Db 331 CACGTCAGCGCATGTGCGCGTGCATCGCGCATCGCGCATCGAGTCCGCGTGCACGCGG 390
Oy 107 SerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeuAlaLeuSerLeuGlyVal 126
Db 391 GCGGATATGACGCGCATCGCTATACCGCGGTCCGCGCTGTGCGCGCTGTGCGTGTG 450
Oy 127 GlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLysProPheIleProIleHis 146
Db 451 GGGGCTTCTGTCGCCAGCGATGCGCTTCCGCTGCGGCGTGCAGCGCGTGCAGCGG 510
Oy 147 HisMetGluAlaHisIleAlaLeuThrIleArgLeuThrAsnLys--ValGluPheProPhe 165
Db 511 CACATGGAAGGCGACCTGCTGGCGCGCATGCTGGAAGAGACGACCGCGCTTCCGCTTC 570
Oy 166 LeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAsp 185
Db 571 GTCGCTTGTGGTTCGCGGCGGTCAACCCAGTGTGTCGCGGTGACCGGTATCGCGCGC 630
Oy 186 PheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAla 205
Db 631 TACCAATTCTTGGCAATCGGTGACGATGCGCGCGCGCAACCTTTCGACAGACCGCC 690
Oy 206 ArgArgLeuSerLeuIleLysIleProGluCysSerThrMetSerGlyGlyAlaIle 225
Db 691 AAGCTATCGGCTG--GGCTATCCC-----GGTGTCCGGAATC 729
Oy 226 GlnHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHis 245
Db 720 GCCCGCTGGCGGACGCGGCACTCTGCGCGCTTGTTCGCGCGCGGATGACCGAT 789
Oy 246 AlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMet 265
Db 790 CGCCCGCGCTGACCTTCAGCTTCAGCGGCGCTCAAGACCTTTACCTTGAAC--ACCTGG 846
Oy 266 LysLysGluLysGlnGlyIleGlyLysGlyGlnIleLeuSerSerAlaAlaAspIle 285
Db 847 CAGCGTTCGCTCAAGCGCGCGACGACGACGACGACG--ACCGCTGCGACATC 897
Oy 286 AlaAlaThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisAlaGlnAla 305
Db 898 GCCCTGGCGTTCAGACCGCGGTGTGACACCTGTGATCAAGTCCGTCGCGCTTG 957
Oy 306 LeuPheCysLysGlnArgAspLeuLeuProGlnAsnAlaValLeuValAlaSerGly 325

```

```

Db 958 -----AAGCAGACCGGCTG-----AAGAAC-----CTGGTATCGCGGC 993
Oy 326 GlyValAlaSerAsnPhetIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGln 345
Db 994 GGTGTAGGCGCAACAGCGCGCTGCGACGCGCTGGAAGAATGCTCGCGCAAAATGAA 1053
Oy 346 CysThrLeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaTrp 365
Db 1054 GGGCAGGTGTTTCAACCCCGCGCGCTTGTGACACGACATGCGCGATGATCGCTTAC 1113
Oy 366 AsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlyIleArg 385
Db 1114 GCCGCTGCCAGCGCGCTGCTCGCGGC-----CAGCATGACGCGCGCGATGACG 1164
Oy 386 TyrGluProLysCysProLeu 392
Db 1165 GTCACGCGCGCTGCGCATG 1185

```

RESULT 9

```

US-09-252-991A-884/C
; Sequence 884, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 884
; LENGTH: 1059
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-884

```

Alignment Scores:

Pred. No.:	1,526-52	Length:	1059
Score:	492.50	Matches:	128
Percent Similarity:	53.24%	Conservative:	61
Best Local Similarity:	36.06%	Mismatches:	143
Query Match:	23.18%	Indels:	23
		Gaps:	9

US-10-649-273-2 (1-414) x US-09-252-991A-884 (1-1059)

```

Oy 39 ValLeuGlyIleGluThrSerCysAspAspThrAlaAlaValAlaAspGluThrGly 58
Db 1038 GTGCTGGAGCTGGAAGCTCTGCGACGAAACCGCGCTGCTTACGACGAAAGCC 979
Oy 59 AsnValLeuGlyGluAlaIleHisSerGlnThrGluValHisLeuLysThrGlyGlyIle 78
Db 978 GGCCTGTGCGCGACGCGGTGTTCAGTCAATCACTTCATCCGCTTACGCGGAGCTC 919
Oy 79 ValProProAlaIleGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGlnAla 98
Db 918 GTGCCCAAGCTGCGCTGCGCGACCATCTCAAGCGCATGCGCGCTGATCGCGCAGGTG 859
Oy 99 LeuSerAlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysPro 118
Db 858 CTCACAGAGTCCGCGTGCACCGCGGATATGACGCGATGCGCTTATACCGCGCGTCC 799
Oy 119 GlyLeuAlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeu 138
Db 798 GGCCTGTGCGCGCGCTGCTGTGCGGCTTCTGTGCGCAGCGCATGCGCTTGCCTGG 739
Oy 139 LysLysProPheIleProIleHisHisMetGluAlaHisIleAlaLeuThrIleArgLeuThr 158

```

```

Db      738 GCGGCGCGCGCGTGGCGCTGACACACATGAAAGGACCTGCTGGCGCGGATGCTGAA 679
Qy      159 AsnlyS---ValGlupheProphleuValleuLeuIleSerglyGlyHisCySleuLeu 177
      678 GAGCAGCACCGCGGTTCCGTCGTCGCTGCTGTTTCCGGCGGTGACACCGATTG 619
Qy      178 AlaIeuValGInglyValSerAspHeuLeuLeuGlyLySleuSerleuAlaPro 197
      618 GTGCGGCGTGAACGGATCGGCGCGCTACACAGTTGCTTGGCAATCGGTGACGATGCGCGC 559
Qy      198 GlyAspMetLeuAspLySValAlaArgArgLeuSerLeuIleLySHisProGluCySer 217
      558 GCGCAAGCTTGTGACAGACCGCGCACTGATCGGCTG---GCTATGCC----- 511
Qy      218 ThrMetSerglyGlyLySAlaIleGluHisLeuAlaLySgInglyAsnArgPheHisPhe 237
      510 -----GGTGTCCGGAATTCGCCGCTGGCGGAGCGCGCACTCCTGCGCGCTTC 460
Qy      238 AspIleLySProProLeuHisHisAlaLySAsnCySAspPheSerPheThrGlyLeuGln 257
      459 GTGTTCCCGCGCGCATGACCGATCGCGCGCTGCGCTTCACTTCAGCTTCAGCGGCTCAAG 400
Qy      258 HisValThrAspLySleuIleMetLySlySgIuLySgIuGlyIleGlyLySgIuGln 277
      399 ACCTTACCTTGAC---ACCTGCGAGCGCTGCGTGCAGCGCGCGCAACGACGACGACG 343
Qy      278 IleLeuSerSeraIAlaAspIleAlaIleThrValGlnHisThrMetAlaCyHisIleu 297
      342 -----ACCGCTGCGACATCCGCCCTGCGGCTTCCAGACCGCGGTGTCAGACCTG 292
Qy      298 ValLySArgThrHisArgAlaIleLeuPheCySlySgIuArgAspLeuProGlnAsn 317
      291 CTGATCAAGTCGCGCGCGCTTG-----AAGCAGACCGCGCTG-----AAGAAC 247
Qy      318 AsnAlaValIleuValAlaSerGlyValAlaIleSerAspPheThrIleArgAlaIleu 337
      246 -----CTGCTATCGCGCGCGGTGTCAGCGCAACGACGCGTGCAGCGCGCTG 196
Db      338 GluIleuThrAsnAlaThrGInCySThrLeuLeuCySProProArgLeuCySThr 357
      195 GAAAGATCTCGCGGAATGAAGGGGACGAGTTCTACGCCCGCGCGCTTCTCACC 136
Qy      358 AspAsnGlyIleMetIleAlaIleArgAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIle 377
      135 GACATGCGCGGATATCGGCTTACCGCGCGCTGCCAGCGCTGCTGCCCGC----- 85
Qy      378 LeuHisAspIleGluGlyIleArgGlyArgProLySAspProleu 392
      84 CAGCATGACGCGCGCGCATGACGCTCCAGCGCGCTGCGCATG 40

```

```

RESULT 10
US-09-543-681A-2341
; Sequence 2341, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; PRIOR FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 2341
; LENGTH: 1074
; TYPE: DNA
; ORGANISM: Proteus mirabilis
US-09-543-681A-2341

```

```

Alignment Scores: 2.97e-51 Length: 1074
Pred. No.: 482.50 Matches: 122
Score:

```

```

Percent Similarity: 49.58%
Best Local Similarity: 34.17%
Query Match: 22.71%
DB: 4 Gaps: 8

US-10-649-273-2 (1-414) x US-09-543-681A-2341 (1-1074)

Qy      39 ValLeuGlyIleGluThrSerCySAspAspThrAlaIleAlaValAlaAspGluThrGly 58
      58 GTTTAGGATTTGAAACATCTTGGCATGAAACCGGATTCGCAATTTCAGATGAAGGCC 117
Db      59 AsnValLeuGlyGluAlaIleHisSerGlnThrGluValHisLeuLySThrGlyIle 78
      118 GGTCTGTAGCGAATCAACTCTATAGCGCAATAAATGACACGCGCACTAAGTGGTGT 177
Qy      79 ValProAlaIleGInGlnLeuHisArgGluAsnIleGlnArgIleValGInGluAla 98
      178 GTTCCGAGCTTGTACAGATCATATCCGAAACAGTGCACATGATCCAAAGCCGCC 237
Qy      99 LeuSerAlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLySPro 118
      238 CTTAAAGAACCAATTAACTGCAAAAGATATGATCGGTCTTATCTGACGACGACA 297
Qy      119 GlyLeuAlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGInLeu 138
      298 GCGCTGATGAGGCGCATGCTGTGGGGGCGACCATGCTGCTTCTGCGGTTGATGG 357
Qy      139 LySLeuProPheIleProIleHisMetGluAlaHisAlaLeuThrIleArgLeuThr 158
      358 GATGTGCTCTATTCAGTTACCATATGAAAGGCAATTTATTAAGCCGAGCTTGAA 417
Qy      159 AsnLySVal---GluPheProPheLeuValleuLeuIleSerglyGlyHisCySleuLeu 177
      418 GAGAAAGCCCAATTTCTTTCGTGCGCTTACTTGTCTCGGGGGGCAATACCAATTA 477
Qy      178 AlaIeuValGInglyValSerAspPheLeuLeuGlyLySleuSerleuAlaIlePro 197
      478 ATTAGCTTAACAGGATGAGGAATATACCTGTAGGTGATGCTATGATGATGCTGT 537
Qy      198 GlyAspMetLeuAspLySValAlaArgArgLeuSerleuIleLySHisProGluCySser 217
      538 GGTGAAGCATTTATTAACAGCAAGCTTTTGGGCTT---GATATGCC----- 585
Qy      218 ThrMetSerglyGlyLySAlaIleGluHisLeuAlaLySgInglyAsnArgPheHisPhe 237
      586 -----GCGCGCGCTGTTATCAAAAATGSCAACAAGGTGTAAGAAAGACGTTT 636
Qy      238 AspIleLySProProLeuHisHisAlaLySAsnCySAspPheSerPheThrGlyLeuGln 257
      637 GTTTTCTGTCGTCATGACAGACAGACCGGACGACCTTATGCTTCTGCTGTTAAA 696
Qy      258 HisValThrAspLySleuIleMetLySlySgIuLySgIuGlyIleGlyLySgIuGln 277
      697 ACCTTCCCGCTAATGCTATTCGTAACAAAGATGATCAGAG----- 728
Qy      278 IleLeuSerSeraIAlaAspIleAlaIleThrValGlnHisThrMetAlaCyHisIleu 297
      739 -----CAACTGACAGATATGCGCGCTTTCGAAGATGCCGTAAGATACCTTGG 792
Qy      298 ValLySArgThrHisArgAlaIleLeuPheCySlySgIuArgAspLeuLeuProGlnAsn 317
      793 GCATTAATAATGTGTCGACG-----TTAGACCAACA 825
Qy      318 Asn---AlaValIeuValAlaSerGlyValAlaIleSerAspPheThrIleArgAla 336
      826 GCGTTAAACGCTTAGATGAGTGGGCGGTAAAGCTAACCGTACTTAAACGCCA 885
Qy      337 LeuGluIleuThrAsnAlaThrGInCySThrLeuLeuCySProProArgLeuCyS 356
      886 ATGCGATGATATATGAAACAATCGAGGGGAAGTGTATTATCTGCCCTGAGATTATGT 945
Qy      357 ThrAspAsnGlyIleMetIleAlaIleArgGlyIleGluArgLeuArgAlaGlyLeu--- 375

```

Db 946 ACCGATATGTCGTCATGTTGGCGGAGATGATCCGTTTAAAGTGTACCGAG 1005
Qy 376 GlyIleuHisAspIleGluGlyIleArgTyrGluProLeuCyseProleu 392
Db 1006 GGGGCTTTA-----GGGTGACAGTGAACACAGTTGGCTTTA 1044
RESULT 11
US-09-902-540-6612
; Sequence 6612, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIORITY DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 6612
; LENGTH: 996
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-6612
Alignment Scores:
Pred. No.: 3,01e-51 Length: 996
Score: 482.00 Matches: 128
Percent Similarity: 52.48% Conservative: 52
Best Local Similarity: 37.32% Mismatches: 131
Query Match: 22.68% Indels: 32
Gaps: 9
US-10-649-273-2 (1-414) x US-09-902-540-6612 (1-996)
Qy 38 IleValIleuGlyIleGluThrSerCyseAspThrAlaAlaValValAspGluThr 57
Db 4 CTCGCTCCTAGAGACTGGAAACCTCGTGTATGAGACTGCCGCCCTCGTGAGGACGGC 63
Qy 58 GlyAsnValIleuGlyGluAlaIleHisSerGlnThrGluValHisLeuYstnGly 77
Db 64 CGCCGCGGTGTGTCATGTCCTCCACGACGTGACATCCACCGCGGTGGG 123
Qy 78 IleValProProAlaIleGlnIleuHisArgGluAsnIleGlnArgIleValGlnGlu 97
Db 124 GTGTCGTCGAGACTGGCCGACCGCAACATCTCCAGGTGTCGCCCTGTCACAG 183
Qy 98 AlaLeuSerAlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleYs 117
Db 184 GCGGTACGCGGCGCAACAGCTCGACGACGTGACCTCATCGCCGTCACGTCCGC 243
Qy 118 ProGlyLeuAlaLeuSerIleuValGlyLeuSerPheSerIleuIleuValGlyGln 137
Db 244 CCCGATCTATCGCGGCTGTGTGTGAGAGTGCAGTGCACAAAGGCTTGACCTGGC 303
Qy 138 LeuYsYsProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeu 157
Db 304 ACGGGCAAGCCCTTCTGTGGGCGCCCAACACCTGAGGGCCACTGTCGCAATCCGGCTG 363
Qy 158 ThrAsnYsVal---GluPheProPheLeuValIleuIleHisSerGlyGlyHisCysLeu 176
Db 364 TTGAGGTGGCGCCGAGCGCGCGCTTCTTGGGCTGTGCTTCCGGCGGACACAGC 423
Qy 177 LeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyYsSerIleuAspIleAla 196
Db 424 CTCTCAGAGTGCAGGCTTACGCGCAGTACCGGCTGTGTGGCGACGCGCGACGAGCGC 483
Qy 197 ProGlyAspMetLeuAspIleValAlaArgArgLeuSerIleuIleHisProGluCys 216
Db 484 GCGGCGAGGATATGACAAAGACCGCTCGCATCTCGGCTG---CCGTATCCG----- 534

Qy 217 SerThrMetSerGlyIleYsAlaIleGluHisLeuAlaYsGlnGlyAsn----- 233
Db 535 -----GGTGGGAGCCCATGACCACTGTCGCGGAGGAAACCGGAGGC 582
Qy 234 ---ArgPheHisPheAspIleYsProProLeuHisIleAlaYsAsnCyseAspPheSer 252
Db 583 ATCCGCTTC-----CCGCGCGCTCCCGGCGCAACATTCGACGTGTCC 627
Qy 253 PheThrGlyLeuGlnHisValThrAspYsIleIleMetYsYsGluYsGluGly 272
Db 628 TTCTCCGGGTTGAG-----ACGCGGTGTGACCACTGTCGCAAGACAGC 675
Qy 273 IleGluYsGlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThr 292
Db 676 GTGCCCGAGGCGGAGCGCGT-----GCCGATTGTGGCCGCTCTCCAGGAGGC 726
Qy 293 MetAlaCysHisLeuValYsArgThrHisArgAlaIleLeuPheCysYsGlnArgAsp 312
Db 727 GTGCGGAGACGTGTCGTAAGAG-----CTGGTGGCCGCGCGCGCG 771
Qy 313 LeuLeuProGlnAsnAsnAlaValIleuValAlaSerGlyYsIleValAlaSerAsnPheTyr 332
Db 772 TTG-----GCCCAAGACAGTTGTGTGTGTGCGCGCGCGCTCGCGGAATCGCG 822
Qy 333 IleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCyseProPro 352
Db 823 CTCGCGGACACTGTGTACGCGCGGACCGGAGCGGCGGTGAACATGTCCTGCCCCG 882
Qy 353 ProArgLeuCyseThrAspAsnGlyIleMetIleAlaIleTrpAsnGlyIleGluArgLeuArg 372
Db 883 GTGCGGTGTGCAACGACATGCGCCATGATGCGGTGCGGGGTATGAGCGTACCGC 942
Qy 373 AlaGlyLeu 375
Db 943 CGCGGCTCG 951
RESULT 12
US-09-902-540-503/C
; Sequence 503, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIORITY DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 503
; LENGTH: 2582
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-503
Alignment Scores:
Pred. No.: 1.62e-50 Length: 2582
Score: 482.00 Matches: 128
Percent Similarity: 52.48% Conservative: 52
Best Local Similarity: 37.32% Mismatches: 131
Query Match: 22.68% Indels: 32
Gaps: 9
US-10-649-273-2 (1-414) x US-09-902-540-503 (1-2582)
Qy 38 IleValIleuGlyIleGluThrSerCyseAspThrAlaAlaValValAspGluThr 57
Db 1525 CTCGTCTTAGAGACTGGAACCTCGTGTATGAGACTGCCGCCCTGTGTGAGAGACGCGC 1466

Qy 58 GlyAsnValLeuGlyValAlaIleHisSerGlnThrGluValHisLeuValThrGlyGly 77
Db 1465 GCGCGCGGCTTCGATGTCCTCCAGAGTGAACATCCAGCGCGGGGCTGG 1406
Qy 78 IleValProProAlaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGly 97
Db 1405 GTGGTCCCGAGACTGGCCAGCCGCAACCATGCTCCAGTGTCCCTCCGTCACGAG 1346
Qy 98 AlaLeuSerAlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLys 117
Db 1345 GCGGTGACGCGCGCGCAACAGACGCTGACGACGACCTGACCTGACCTGACCTGCGG 1286
Qy 118 ProGlyLeuAlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGln 137
Db 1285 CCGGACTCATCTGCGCGCTGCTGTGGAGTGCAGTGCAGGCGCTGACCTGCGCG 1226
Qy 138 LeuLysLysProPheIleProIleHisMetGluAlaHisAlaLeuThrIleArgLeu 157
Db 1225 ACGGCGAAGCCCTTGCGGGCGCCCAACCTGAGAGGCGCACCTGCGCATCCGCGTGG 1166
Qy 158 ThrAsnLysVal---GluPheProPheLeuValIleLeuIleSerGlyHisCysLeu 176
Db 1165 TTGAGAGTGGCGCGCGAGCGCGCTTCCTTGCGCTGCTGCTTCGCGCGGACACACAG 1106
Qy 177 LeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIleAla 196
Db 1105 CTCTACGAGTGCAGGCTACCGGCGAGTACCGGCTGCTGGGCGACGACGCGCGAGCGG 1046
Qy 197 ProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGluCys 216
Db 1045 GCGCGCGAGGATATGACAAAGACCGCTCCATCTCCGCGCTG---CCGATCCG----- 995
Qy 217 SerThrMetSerGlyValAlaIleGlnHisLeuAlaLysGlnLysAsn----- 233
Db 994 -----GTTGGGACGCCCATGACCACTTGCGCGAGGAGGAGGAGGAGGCGC 947
Qy 234 ---ArgPheHisPheAspIleLysProPheLeuHisAlaLysAsnCysAspPheSer 252
Db 946 ATCCGCTTC-----CCGCGCGCGCTGCGCGGCGACAACTTCGACGCTGTC 902
Qy 253 PheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLysGlnLysGlnGly 272
Db 901 TTCTCCGCGCTGAG-----ACGCGCGGCTCTCACCACTGACGAGACGCGC 854
Qy 273 IleGluLysGlyGlnIleLeuSerSerAlaIleAspIleAlaAlaThrValGlnHisThr 292
Db 853 GTGCGCGAGGCGCGCGCTG-----GCGGATTTGTGCGCGCTTCCTCCAGAGGCC 803
Qy 293 MetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAsp 312
Db 802 GTGCGGACGTCCTCGAAGAG-----CTGCTGCGCGCGCGCGCGCGCG 758
Qy 313 LeuLeuProGlnAsnAlaValLeuValAlaSerGlyValAlaSerAsnPheThr 332
Db 757 TTG-----GCGCACAGCAGTTGTGTGTGCGCGCGCGCTCCCGCAACTCGCGG 707
Qy 333 IleArgGalaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProPro 352
Db 706 CTGCGGCGACTGTCTGACGCGGAGCGGAGCGGCGGTGAACTGTTCCTGCGCGCG 647
Qy 353 ProArgLeuCysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArg 372
Db 646 GTGCGGCTGTGACGAGCAATGCGCGCATGATTCGCGTGGCGGGATATAGCGCTACCG 587
Qy 373 AlaGlyLeu 375
Db 586 GCGGCGCTG 578

RESULT 13
US-08-087-797-1
; Sequence 1, Application US/08087797
; Patent No. 5543312
; GENERAL INFORMATION:

APPLICANT: Mellors, Alan
APPLICANT: Lo, Reggie Y.C.
APPLICANT: Abdullah, Khalid M.
TITLE OF INVENTION: Pasteurella Haemolytica
TITLE OF INVENTION: Glycoprotease
TITLE OF INVENTION: Gene and the Purified Enzyme
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Bell, Seltzer, Park & Gibson, P.A.
STREET: 1211 East Morehead Street,
CITY: Charlotte
STATE: No. 5543312ch Carolina
COUNTRY: United States
ZIP: 28234
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/087,797
FILING DATE: 14-JUL-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Layton, Jr., Samuel G.
REGISTRATION NUMBER: 22807
REFERENCE/DOCKET NUMBER: 3374-80
TELECOMMUNICATION INFORMATION:
TELEPHONE: 704 377 1561
TELEFAX: 704 334 2014
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1315 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 141..1115
US-08-087-797-1
Alignment Scores:
Pred. No.: 1.19e-50 length: 1315
Score: 479.00 Matches: 128
Percent Similarity: 48.63% Conservative: 50
Best Local Similarity: 34.97% Mismatches: 157
Query Match: 22.54% Indels: 31
DB: 1 Gaps: 8
US-10-649-273-2 (1-414) x US-08-087-797-1 (1-1315)
Qy 9 GlyValPhePheLysProSerLysArgLysValIleGluPheLeuArgSerPheAsp 28
Db 89 GGGGCGTTCTTCCCGCT-----TTTGCTTTCTACTTATTTGACTTC 133
Qy 29 HisProGlyThrLeuPheLeuHisLysIle-ValLeuGlyIleGluThrSerCysAspAs 48
Db 134 TCC-----AACTATGCAATTTATGATTAATAAAGCTCTTGATGTA 175
Qy 48 pThrAlaAlaValAlaValAlaValAlaValAlaValAlaValAlaValAlaValAla 68
Db 176 AACCGGTGTGCGCATTTATGATGAGCAAGAGCTTGTGCGCAACGCTTATATACCA 235
Qy 68 nThrGluValHisLeuLysThrGlyGlyIleValIleProProAlaIleGlnGlnLeuHis 88
Db 236 AATTGATATTCAGCGCGCATTTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 295
Qy 88 gGluAsnIleGlnArgIleValGlnGlnAlaLeuSerAlaSerGlyValSerProSerAs 108
Db 296 CGCTAAAGCTTCCCACTAATTCAAGAGCTTAAAGAGGCCAATCTGACACCTCGGA 355

QY 108 pLeuSerAlaIlealThrThrIleLeuSProGlyLeuAlaLeuSerLeuGlyValGlyLe 128
Db 356 TATTGACGGCATTTGCTTACTATGCGGCGCCAGGCTTGCGGCTTATATGCTGCGCTC 415
QY 128 uSerPheSerLeuGlnLeuValGlyGlnLeuLysPProPheIleProIleHisHisIle 148
Db 416 AACCATTTGCCGTTTCCGCTTATGCTTGAATGTTCCGCAATGGGCTTACCATAT 475
QY 148 tGluAlaHisAlaLeuThrIleArgLeu---ThrAsnLysValGluPheProPheLeu 167
Db 476 GGAAGGGCATTTACTTGCCCAATGTGGAAGAAAATGCCCCCTGAATTTCCGTTTGCGC 535
QY 167 lLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPhe 187
Db 536 ATTATTTGATTTTCAAGTGGACACACCACTGTAAAGTTGAACGGGTTGGCAATACGA 595
QY 187 uLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgAr 207
Db 596 ACTACTCGGGGAATCAATGATGATGATGCTGCCGTTGAAGCTTTGACAAACAGGCAAACT 655
QY 207 gLeuSerLeuIleLysPheSProGlyCysSerThrMetSerGlyGlyValAlaIleGlu 227
Db 656 ACTCGTTTG---GATTACCTT-----GCCGGTGAAGCATGTCAA 694
QY 227 sLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaL 247
Db 695 ATTAGCCGAATCCGACCGCCAAATCGTTTAAATTCCTCGTCCAAATGACGACGACG 754
QY 247 sAsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLys 267
Db 755 GGGACTGATTTTCAAGTTCTCCGTTTAAACCTTGTGCGCAATACATTAAGCCAA 814
QY 267 sGluLysGluGluGlyIleGlnLysGlyGlnIleLeuSerSerAlaAlaAspIleAla 287
Db 815 TCTTAATGAAATGCTGAATCTGATGAGCAA-----ACCAATGCGATTTGCCCA 865
QY 287 aThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeu 307
Db 866 CGCATTCACAAACCCGTC-----GTTGATACATTTTAAATTAA 904
QY 307 eCysLysGlnArgAspLeuLeuProGlnAsnAlaValLeuValAlaSerGlyVal 327
Db 905 ATGCAAG---CGAGGCTTACAGCAACCGGCTTAAACGCTTATGATGACGCGCGCT 961
QY 327 lAlaSerAspPheTyrlleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCys 347
Db 962 AAGTCCCAATAAACAATTACGAGCAGACCTTGCGGAATGATGAAAAAATTAAAGCGCA 1021
QY 347 rLeuLeuCysProProArgLeuCysThrAspAsnGlyIleMetIleAlaTrpAsn 367
Db 1022 AGTATCTTCACTCGCCCAAAATTTTGCACTGACAAAGGCGCAATGATGCTTACTG 1081
QY 367 yIleGluArgLeuArg 372
Db 1082 CTTTCTTCCTTAA 1097

RESULT 14
US-09-557-884-1/c
; Sequence 1, Application US/09557884
; Patent No. 6506581
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: The Nucleotide sequence of
; the Haemophilus influenzae Rd Genome, Fragments
; Thereof, and Uses Thereof
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: MD
; COUNTRY: USA

ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: 3 1/2 inch diskette
COMPUTER: Dell Pentium
OPERATING SYSTEM: MS DOS v6.22
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/557,884
FILING DATE: 25-Apr-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/476,102
FILING DATE: JUN-5-1995
ATTORNEY/AGENT INFORMATION:
NAME: Michelle S. Marks
REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: PB186P3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-309-8504
TELEFAX: 301-309-8439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1830121 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1
Alignment Scores:
Pred. No.: 1.21e-44 Length: 1830121
Score: 475.50 Matches: 120
Percent Similarity: 50.74% Conservative: 52
Best Local Similarity: 35.40% Mismatches: 144
Query Match: 22.38% Indels: 23
DB: 4 Gaps: 7
US-10-649-273-2 (1-414) x US-09-557-884-1 (1-1830121)
QY 39 ValLeuGlyIleGluThrSerCysAspAspThrAlaAlaValAlaAspGluThrGly 58
Db 553208 ATCTTAGCGCATTAATAAATCTCTGATGAAACGGCGCTGCGATTTGATGAAAGAAA 553149
QY 59 AsnValLeuGlyGluAlaIleHisSerGlnThrGluValHisLeuLysThrGlyIle 78
Db 553148 GGAATTAATCTTATGATCACTTATATCAATGATGCGCTGATGACGATTAAGTGATG 553089
QY 79 ValProProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGlu 98
Db 553088 GTTCTGAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 553029
QY 99 LeuSerAlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleAsp 118
Db 553028 TTAGAGGAGCCCAATTAACCGGAGGAGATGATGATGATGATGATGATGATGATGAT 552969
QY 119 GlyLeuAlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGln 138
Db 552968 GGGCTTTTGGCGCATTTGCTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 552909
QY 139 LysLeuProPheIleProIleHisHisMetGluAlaHisAlaLeuThrIleArgLeu 157
Db 552908 AATGTTCTCGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 552849
QY 158 ThrAsnLysValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeu 177
Db 552848 GACAAATTCACCGCATTTCTTTGTTGCTCTGTTGATGCGGAGGACACTCAATTA 552789
QY 178 AlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIleAlaPro 197
Db 552788 GTGCGTGTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 552729
QY 198 GlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysPheProGlyCysSer 217

Db 552728 GCGAAGCCTTGATTAACAGCAAAATTACTTGACTA---GATTATCCA----- 552681

Qy 218 ThrmcSerGlyGlyValSalIleGluHisLeuAlaLysGlnGlyAsnArgPheHisPhe 237

Db 552680 -----GGTGGCGCGGACATTTCTCGTTAGCGGAAAAAGGTACGCCAAATCGTTTC 552630

Qy 238 AspIleLysProProLeuHisHisSalAlaLysAsnCyAspPheSerPheThrGlyLeuGln 257

Db 552629 ACATTTCACGTCGATGACAGATCGTCGAGCGCTTGATTTTGTGTTTAAATA 552570

Qy 258 HisValThrAspLysIleIleMetLysGlyLysGlnGluGlnGly-----11GluLys 275

Db 552569 ACATTGGCCGCAATACATTAATCAAGCAATTAATAACAGAGCGCAATGATGAGCAA 552510

Qy 276 GlyGlnIleLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCys 295

Db 552509 -----ACTAAGAGATATTGCTTATGCTTCCAAAGATGGCGGTGGAT 552465

Qy 296 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 315

Db 552464 ACTCTTGCC-----ATTAATGTAG--CGTCATTGAAAGAA 552429

Qy 316 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheThrIleArgArg 335

Db 552428 ACAGGCTATAACGTTTACGTGATTGCGGAGGGGTAGCGCAATTAATAAACTCCGAGAA 552369

Qy 336 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLysCysProProArgLeu 355

Db 552368 ACGCTTGCCACATTAATGCAAAATTAGGTGCGGAAGTGTATTATCTCAACCTCAATTT 552309

Qy 356 CysThrAspAsnGlyIleMetIleAlaIleArgAsnGlyIleGluArgLeuArgAlaGly 374

Db 552308 TGTACAGATTAATGTCGATGATGCTTACACAGGTTTTTTTACGTTTAAACAAGGT 552252

RESULT 15

US-09-643-990A-1/c

Sequence 1, Application US/09643990A

Patent No. 6528289

GENERAL INFORMATION:

APPLICANT: Robert D. Fleischmann

Mark D. Adams

Owen White

Hamilton O. Smith

J. Craig Venter

TITLE OF INVENTION: The Nucleotide sequence of

The Haemophilus influenzae Rd Genome, Fragments

Thereof, and Uses Thereof

NUMBER OF SEQUENCES: 1

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville,

STATE: MD

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: 3 1/2 inch diskette

COMPUTER: Dell Pentium

OPERATING SYSTEM: MS DOS v6.22

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/643,990A

FILING DATE: 23-Aug-2000

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/487,429

FILING DATE: 1995-06-07

APPLICATION NUMBER: 08/426,787

FILING DATE: 1995-04-21

ATTORNEY/AGENT INFORMATION:

NAME: Kenley K. Hoover

REGISTRATION NUMBER: 40,302

REFERENCE/DOCKET NUMBER: PB186P1C1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 301-610-5790

TELEFAX: 310-309-8439

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1830121 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-643-990A-1

Alignment Scores:

Pred. No.: 1,21e-44 Length: 1830121

Score: 475.50 Matches: 120

Percent Similarity: 50.74% Conservative: 52

Best Local Similarity: 35.40% Mismatches: 144

Query Match: 22.38% Indels: 23

DB: 4 Gaps: 7

US-10-649-273-2 (1-414) x US-09-643-990A-1 (1-1830121)

Qy 39 ValLeuGlyIleGluThrSerCysAspAspThrAlaAlaValAlaAspGluThrGly 58

Db 553208 ATCTTAGGCAATGAACTTCTGTGATGAAACCGGCGTGGCATTTAGTGAAGAAAA 553149

Qy 59 AsnValLeuGlyGluAlaIleHisSerGlnThrGluValHisLeuLysThrGlyIle 78

Db 553148 GGAATTAATGCTTAATCACTTATATCTCAAAATGCCCTGCATGCAATTAATGTGTGTG 553089

Qy 79 ValProProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAla 98

Db 553088 GTTCTGAATTAGATGATCAGCGATCATATTCGAAACAGACCTCTTATTAACGGCA 553029

Qy 99 LeuSerAlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysPro 118

Db 553028 TTAGAGGAGGCCAATTTAAACCGGACGATATGATGATGCTTATACGATGGCCCT 552969

Qy 119 GlyLeuAlaLeuSerLeuGlyValGlyLysSerPheSerLeuGlnLeuValGlyGlnLeu 138

Db 552968 GGGCTGTGGCGCATGCTTGTCGTGCTACATTCACGCTTATAGCCTATGCTTGG 552909

Qy 139 LysLysProProIlePheProIleHisMetGluAlaHisAlaLeuThrIleArgLeu--- 157

Db 552908 AATGTTCTCGCATGCTGCTTATCATATGAAAGGCAATTAATCTGGCCCAATGCTGAT 552849

Qy 158 ThrAsnLysValGluPheProPheLeuValLeuLeuIleSerGlyGlnHisCysLeuLeu 177

Db 552848 GACAAATTCACCGCACCTTCTTTGTTGCTGTGGTATCGGGTGCACACATCAATTA 552789

Qy 178 AlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIleAlaPro 197

Db 552788 GTGCGTGTGATGATGATGAAATATGAAGTATGAGTAATCTATTTGATGTGCTGT 552729

Qy 198 GlyAspMetLeuAspLysValAlaArgLysSerLeuLeuLysHisProGluCysSer 217

Db 552728 GGGAAAGCCTTATTAATAACAGAAATTAATCTTGACTA---GATTATCCA----- 552681

Qy 218 ThrmcSerGlyGlyValSalIleGluHisLeuAlaLysGlnGlyAsnArgPheHisPhe 237

Db 552680 -----GGTGGCGCGGACATTTCTCGTTAGCGGAAAAAGGTACGCCAAATCGTTTC 552630

Qy 238 AspIleLysProProLeuHisHisSalAlaLysAsnCyAspPheSerPheThrGlyLeuGln 257

Db 552629 ACATTTCACGTCGATGACAGATCGTCGAGCGCTTGATTTTGTGTTTAAATA 552570

Qy 258 HisValThrAspLysIleIleMetLysGlyLysGlnGluGlnGly-----11GluLys 275

Db 552569 ACATTGGCCGCAATACATTAATCAAGCAATTAATAACAGAGCGCAATGATGAGCAA 552510

Qy 276 GlyGlnIleLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCys 295

Db 552509 -----ACTAAGAGATATTGCTTATGCTTCCAAAGATGGCGGTGGAT 552465

```

Qy      296 HisLeuValIysArgThrHisArgAlaIleuPheCysLysGlnArgAspLeuPro 315
Db      552464 ACTCTTGCC-----ATTAATGTAAG---CGTCATTGAAAGAA 552429
Qy      316 GlnAsnAsnAlaValIleuValAlaSerGlyIValAlaSerAsnPheTyrIleArgArg 335
Db      552428 ACAGCCTATAAAGCTTGTAGTGTGGGAGGGGTGACCGCAATATAAAACTCCGAGAA 552369
Qy      336 AlaLeuGluIleuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 355
Db      552368 ACGCTTGCGCACTTAATGCAGAAATTAGGTGGCGAAGTGTATTATCTCAACCTCAATTT 552309
Qy      356 CysThrAspAsnGlyIleuEtilleAlaTpaAsnGlyIleGluArgLeuArgAlaGly 374
Db      552308 TGTACAGATTAATGTGTGCGATGATTGCTTACACAGGTTTTTTACGTTTAAACAGAGGT 552252

```

Search completed: November 11, 2005, 02:47:48
 Job time : 1671.85 secs


```

Db 471 CCAGACGCTCAACAGCTTCAAGAGAAATATTCAGCATAGTACAGAAAGCTCTTCT 530
Qy 101 AlAsereglValSerProSerAspLeuSerAlaIlealThrThrIleLeuProglYleu 120
Db 531 GCCAGTGGAGCTCTCCCAAGTAGACCTCTCAGCAATGTGACATACATAAACAGAGACTT 590
Qy 121 AlaleuSerleuGlYValleuSerPheSerleuGlYleuValGlYleuYslys 140
Db 591 GCTTAAAGCTGGAGAGGCTTATTCATTACCTTACGCTGGTGGAGACATTAATAAANG 650
Qy 141 ProPheIleProIleHisIleMetGluAlaHisIleuThrIleArgLeuThrAsnYls 160
Db 651 CCATTCAATCCCATTCATCATATGAGAGCTCATGACCTACTATAGGTGACCATATAA 710
Qy 161 ValGluPheProPheLeuValleuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180
Db 711 GTAGAAATTTCTTTTGTAGTTCTTTGTAGTTCTGAGAGCTCATCTGCTGTGGCATATAGT 770
Qy 181 GlnglyValSerAspPheLeuLeuLeuGlyYsSerleuAspIleAlaProglYAspMet 200
Db 771 CAAGAGATTTCAAGATTTCTGCTCTTGAAGCTTTGAGACATAGCACAGAGTACATG 830
Qy 201 LeuAspYsValAlaArgArgLeuSerleuIleYsHisProglYCysSerThrMetSer 220
Db 831 CTTGACAGGTGGCAGAGAGACTTTCTTAATTAACATCCAGAGTCTCCACATGAGT 890
Qy 221 GlyGlyYsAlaIleGluHisleuAlaYsGlnGlyAsnArgPheHisPheAspIleYs 240
Db 891 GGTGGAGAAAGCATGAGACATTTGGCCAAACAGAAATAGATTCATTTGACATCAAA 950
Qy 241 ProProLeuHisHisAlaYsAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260
Db 951 CCTCCCTTGCATCAAGCTAAATATGATGATTTCTTTTACATGAGCTTCAACACGTTACT 1010
Qy 261 AspYsIleIleMetYsLeuGluYsGlnGlyGlyIleGluYsGlyGlnIleLeuSer 280
Db 1011 GATTAATAATATATATAAAGAGAAAGAGAGATTTAGAGAGGGCAATCTCTCT 1070
Qy 281 SerAlaAlaAspIleAlaIleThrValGlnHisIleMetAlaCysHisleuValYsArg 300
Db 1071 TCAGAGAGAGACATTCCTGCCACAGTACAGACACATGAGCATGTCATCTTGAGAAAG 1130
Qy 301 ThrHisArgAlaIleLeuPheCysYsGlnArgAspLeuLeuProGlnAsnAlaVal 320
Db 1131 ACACATCCGGCTATTCGTGTTGTGAAGCAGAGAGACTTCTTAACCTCAAAATATATGACAT 1190
Qy 321 LeuValAlaSerGlyValAlaSerAsnPheYrIleArgArgAlaLeuGlnIleLeu 340
Db 1191 CTGTTGATCTGGGTGGTGGTCCAGATATCTCTATATCCGAGAGCTCTGGAAATTTTA 1250
Qy 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeuCysThrAspAsnGly 360
Db 1251 ACAACGCAACACAGTGCATCTTGTGTCTCTCCCTCCAGACTAGCATGATAATGG 1310
Qy 361 IleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380
Db 1311 ATTATGATTCAGTGAATGATGTAATGAAGACTACGCTGCGCTGGCGCATTTTACATGAC 1370
Qy 381 IleGluGlyIleArgTrpGluProYsCysProLeuGlyValAspIleSerYsGluVal 400
Db 1371 ATAGAGAGGATCCGCTATAGAACCAAAATGTCTCTTGAGGTAGACATATCAAAAGAGTT 1430
Qy 401 GlyGluAlaSerIleYsValProGlnLeuYsMetGluIle 414
Db 1431 GGAGAGGCTTCCATTAAGTACCAATTAATAATGAGATTA 1472

```

RESULT 2

```

US-10-649-273-1
; Sequence 1, Application US/10649273
; Publication No. US20040043407A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1

```

```

; FILE REFERENCE: D0073 CNT
; CURRENT APPLICATION NUMBER: US/10/649,273
; PRIOR FILING DATE: 2003-08-27
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 2197
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (231) .. (1472)
US-10-649-273-1
Alignment Scores:
Pred. No.: 0 Length: 2197
Score: 414.00 Matches: 414
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0
DB: 19
US-10-649-273-2 (1-414) x US-10-649-273-1 (1-2197)
Qy 1 MetLeuIleLeuThrIleYsThrAlaGlyValPhePheYsProSerYsArgYsValYr 20
Db 231 ATGCTATCTTGACTAGTACGAGAGCTTTTAAACCAACAAAGAAAGTTAT 290
Qy 21 GluPheLeuArgSerPheAsnPheHisProglYThrLeuPheLeuHisYsIleValLeu 40
Db 291 GAATTTTAAAGATTTTATTTCACTCGAAGACATATTTCTTAATAATAGTATG 350
Qy 41 GlyIleGluThrSerCysAspAspThrAlaAlaValAlaValAspGluThrGlyAsnVal 60
Db 351 GGAATTAAGAACTGATGATGATGATGATGATGATGATGATGATGATGATGATG 410
Qy 61 LeuGlyGluAlaIleHisSerGlnThrGluValHisleuYsThrGlyIleValPro 80
Db 411 TTGGAGAGACATATATCTCCAACTGAGCTTATTAATAACAGGTGGATTTCT 470
Qy 81 ProAlaAlaGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100
Db 471 CCAGAGCTCAACAGCTTCAAGAGAAATATTCAGATATGTAACAAGAGCTTTCT 530
Qy 101 AlAsereglYValSerProSerAspLeuSerAlaIlealThrThrIleYsProglYleu 120
Db 531 GCCAGTGGAGCTCTCCAGATGACCTCTCAGCAATGCACTCAATTAACAGAGACT 590
Qy 121 AlaleuSerleuGlYValleuSerPheSerleuGlYleuValGlYGlnLeuYslys 140
Db 591 GCTTAAAGCTGGAGAGGCTTATTCATTACCTTACGCTGGTGGAGACATTAATAAAG 650
Qy 141 ProPheIleProIleHisIleMetGluAlaHisIleuThrIleArgLeuThrAsnYls 160
Db 651 CCATTCAATCCCATTCATCATATGAGAGCTCATGACCTACTATAGGTGACCATATAA 710
Qy 161 ValGluPheProPheLeuValleuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180
Db 711 GTAGAAATTTCTTTTGTAGTTCTTTGTAGTTCTGAGAGCTCATCTGCTGTGGCATATAGT 770
Qy 181 GlnglyValSerAspPheLeuLeuLeuGlyYsSerleuAspIleAlaProglYAspMet 200
Db 771 CAAGAGATTTCAAGATTTCTGCTCTTGAAGCTTTGAGACATAGCACAGAGTACATG 830
Qy 201 LeuAspYsValAlaArgArgLeuSerleuIleYsHisProglYCysSerThrMetSer 220
Db 831 CTTGACAGGTGGCAGAGAGCTTTCTTAATTAACATCCAGAGTCTCCACATGAGT 890

```

QY 221 GlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLys 240
DB 891 GGTGGAAAGCCATGAAACATTGGCCAAAGCAAGAAATGATTCATTGATCAATCAAA 950
QY 241 ProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260
DB 951 CTTCCCTTCATCAATGCTAAATAATGATTTCTTTTACTGAGACTTCAACACGTTACT 1010
QY 261 AspLysIleIleMetLysLysGlnLysGlnGlyIleGlnLysGlnIleLeuSer 280
DB 1011 GATTAATAATATATATAAAGAAAGAAAGAGATTTGAGAAGGGCAAAATCTGTCT 1070
QY 281 SerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCysHisLeuValLysArg 300
DB 1071 TCACAGCAGACGATTCGTCCACAGTACAGACACAAATGGCATGTCATCTTGTAAGAAGA 1130
QY 301 ThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuProGlnAsnAlaVal 320
DB 1131 ACACATCGGGCTATTCTGTTTGTGAAGCAGAGACTGTTTACTCAAAATATATCAGTA 1190
QY 321 LeuValAlaSerGlyGlyValAlaSerAsnPheTyrIleArgArgAlaLeuGlnIleLeu 340
DB 1191 CTGGTTGCATCTGGTGGTGTCCAGATTAATCTTATTCGCAAGCTCTGGAATTTTA 1250
QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeuCysThrAspAsnGly 360
DB 1251 ACAACGCAACACAGTCGACTTGTGTCTCTCCACAGACTATGACATATATGAC 1310
QY 361 IleMetIleAlaTyrAsnGlyIleGlnArgLeuAlaGlyLeuGlyIleLeuHisAsp 380
DB 1311 ATATATGATTCATGATGAATGATTAATGAAGACTACCTGCTGGCTGGCTTTTACATGAC 1370
QY 381 IleGlnGlyIleArgTyrGluProLysCysProLeuGlyValAspIleSerLysGlnVal 400
DB 1371 ATTAAGGCAATCCGCTATGAACCAAAATGCTCTTGGAGTGAATCAATCAAAAGAGTT 1430
QY 401 GlyGlnAlaSerIleLysValProGlnLeuLysMetGlnIle 414
DB 1431 GGAGAAGCTTCATTAATAAGTACCAATTAATAAATGAGATA 1472

RESULT 3
US-10-651-722-1
; Sequence 1, Application US/10651722
; Publication No. US20040048302A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1
; FILE REFERENCE: D0073 DIV
; CURRENT APPLICATION NUMBER: US/10/651,722
; PRIOR FILING DATE: 2003-08-29
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: Patent version 3.2
; SEQ ID NO 1
; LENGTH: 2197
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURES:
; NAME/KEY: CDS
; LOCATION: (231)..(1472)
US-10-651-722-1

Alignment Scores:
Preg. No.: 0 Length: 2197
Score: 414.00 Matches: 414
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0

Query Match: 100.00% Indels: 0
DB: 19 Gaps: 0
US-10-649-273-2 (1-414) x US-10-651-722-1 (1-2197)

QY 1 MetLeuIleLeuThrThrIleAlaGlyValPhePheLysProSerLysArgLysValTyr 20
DB 231 ATGCTATATCTGTGCTAAGACTGACAGAGATTTTTTTTAAACATCAAAAGAAAGTTAT 290
QY 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisLysIleValLeu 40
DB 291 GAATTTTAAAGAGTTTAAATTTTCAATCCGGAACATAATTTCTTCAATAAATAGTATG 350
QY 41 GlyIleGluThrSerCysAspAspThrAlaAlaValValAspGluThrGlyAsnVal 60
DB 351 GGATTTGAACATGATGTTGATGATGATACAGAGCTGCTGTGATGAAACTGGAATATG 410
QY 61 LeuGlyGlnAlaIleHisSerGlnThrGluValHisLeuLysThrGlyIleValPro 80
DB 411 TTGGGGAAGCAATATACATTCCTCAACCTGAAGTCAATTAATAAAGAGGATGTTCT 470
QY 81 ProAlaIleGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGlnAlaLeuSer 100
DB 471 CCAGCAGCTCAACAGCTTCAACAGAAATATTCACGAATGATCAAGAAAGCTCTTCT 530
QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeu 120
DB 531 GCCAGTGAAGTCTCTCAAGTACCTCTCAGCAATTCGCAATACATTAACCAAGACCT 590
QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyLeuLysLys 140
DB 591 GCTTTAAGCTGGAGAGTGGCTTATCATTTAGCTTACAGCTGATGAGCAGTTAAAAAG 650
QY 141 ProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLys 160
DB 651 CCATTCATTCCTCATTCATATGATGAGGCTCAGACCTTATCTTATGAGTTGACCAATAA 710
QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180
DB 711 GTAGATTTCTCTTTTATGATCTTTGATTTCTGGAGGTCACCTGTGTTGGCATTAAGT 770
QY 181 GlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet 200
DB 771 CAAGAGTTTCAAGATTTTCTGCTTCTTGAAGATCTTTGACATACACAGGTGACATG 830
QY 201 LeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGluCysSerThrMetSer 220
DB 831 CTGACAAAGTGGCAAGAACTTTCTTAAATTAACATCCAGAGTCTCCACATGAGT 890
QY 221 GlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLys 240
DB 891 GGTGGAAAGCCATGAAACATTGGCCAAAGCAAGAAATGATTCATTGATCAATCAAA 950
QY 241 ProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260
DB 951 CTTCCCTTCATCAATGCTAAATAATGATTTCTTTTACTGAGACTTCAACACGTTACT 1010
QY 261 AspLysIleIleMetLysLysGlnLysGlnGlyIleGlnLysGlnIleLeuSer 280
DB 1011 GATTAATAATATATATAAAGAAAGAAAGAGATTTGAGAAGGGCAAAATCTGTCT 1070
QY 281 SerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCysHisLeuValLysArg 300
DB 1071 TCAGCAGCAGACATTCGTCCACAGTACAGACACAAATGGCATGTCATCTTGTAAGAAGA 1130
QY 301 ThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuProGlnAsnAlaVal 320
DB 1131 ACACATCGGGCTATTCTGTTTGTGAAGCAGAGACTGTTTACTCAAAATATATCAGTA 1190
QY 321 LeuValAlaSerGlyGlyValAlaSerAsnPheTyrIleArgArgAlaLeuGlnIleLeu 340
DB 1191 CTGGTTGCATCTGGTGGTGTCCAGATTAATCTTATATCCGCAAGCTCTGGAATTTTA 1250

Qy 341 ThrAsnaIaThrGlnCysThrIleuLeuCysProProArgLeuCysThrAspaSngly 360
Db 1251 ACAAGCGAACAACATGCACTTGTGTCTCTCCAGCATGCACTATATATGCG 1310
Qy 361 IImecIlealatrPaSnglyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisaP 380
Db 1311 ATTATGATTCGATGAAATGATGAAAGACTACGCTGCTGGGCACTTTTACATGAC 1370
Qy 381 IleguIlyIleArgTyGluProLysCysProLeuGlyValAspIleSerIyGluVal 400
Db 1371 ATAGAGGCACTCCGCTATGAAACCAAAATGCTCTTGAGTACATATCAAAAGAGTT 1430
Qy 401 GlyIuAlaSerIleLysValaProGlnLeuLysMetGluIle 414
Db 1431 GGAGAGCTCCATAAAGTACCAATTTAAAAATGAGATA 1472

RESULT 4
US-10-480-988-36
; Sequence 36, Application US/10480988
; Publication No. US20050069877A1
; GENERAL INFORMATION:
; APPLICANT: GANDHI, Ameena R.; KABLE, Amy E.;
; APPLICANT: SMARNAKAR, Anita; HAFALIA, April J.A.;
; APPLICANT: TRAN, Bao; DUGGAN, Brendan M.;
; APPLICANT: WARREN, Bridget A.; ISON, Craig H.;
; APPLICANT: HONCHELL, Cynthia D.; NGUYEN, Daniel B.;
; APPLICANT: LU, Dyanh Aina M.; LEE, Ernestine A.;
; APPLICANT: YUE, Henry; FORSYTHE, Ian J.;
; APPLICANT: BARROSO, Ines; RAMKUMAR, Jayalaxmi;
; APPLICANT: GRIFFIN, Jennifer A.; LI, Jiana X.;
; APPLICANT: YANG, Junming; THANHAVEU, Kavitha;
; APPLICANT: GIETZEN, Kimberly J.; DING, Li;
; APPLICANT: BAUGHN, Mariah R.; BOROWSKY, Mark L.;
; APPLICANT: YAO, Monique G.; CHAMUA, Nandinder K.;
; APPLICANT: MASON, Patricia M.; GURURAJAN, Rajagopal;
; APPLICANT: LEE, Sally; BECHA, Shanya D.;
; APPLICANT: LEE, Soo Yeun; TRAN, Uyen K.;
; APPLICANT: ELIOTY, Vicki S.; LUO, Wen;
; APPLICANT: SPRAGUE, William W.; TANG, Y. Tom;
; APPLICANT: LU, Yan; ZEBARJADIAN, Yeganeh
; TITLE OF INVENTION: PROTEIN MODIFICATION AND MAINTENANCE MOLECULES
; FILE REFERENCE: PF-1040 USN
; CURRENT APPLICATION NUMBER: US/10/480,988
; CURRENT FILING DATE: 2003-12-16
; PRIOR APPLICATION NUMBER: PCT/US02/19360
; PRIOR FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: US 60/300,508
; PRIOR FILING DATE: 2001-06-22
; PRIOR APPLICATION NUMBER: US 60/303,445
; PRIOR FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US 60/305,405
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/311,442
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: US 60/314,821
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/315,992
; PRIOR FILING DATE: 2001-08-29
; PRIOR APPLICATION NUMBER: US 60/378,205
; PRIOR FILING DATE: 2002-05-03
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: PERL Program
; SEQ ID NO 36
; LENGTH: 2572
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: 7632424CB1
US-10-480-988-36

Alignment Scores: 0 Length: 2572
Pred. No.: 3

Score: 414.00 Matches: 414
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 22 Gaps: 0

US-10-649-273-2 (1-414) x US-10-480-988-36 (1-2572)

Qy 1 MetIleuIleLeuThrIlyThrAlaGlyValaPhePheLysProSerIySArgLysValaIlyr 20
Db 144 ATGCTATATCTTGACTTAAGACTGAGAGATTTTTTTTAAACCATCAAAAAGAAAGTTAT 203
Qy 21 GluPheLeuArgSerPheAsnPhenIspProGlyThrLeuPheLeuHisaIyValleu 40
Db 204 GAATTTTAAAGAGTTTATTTTATCTCTGGAACATATTTCTCATAAATATGATTTG 263
Qy 41 GlyIleGluThrSerCysAspAspThrAlaAlaValaValaAspGluThrGlyAsnVal 60
Db 264 GGAAATTGAACCTAGTTGTGATGATACAGCAGCTGCTGTGTGATGAAACCTGAAATGTG 323
Qy 61 LeuGlyGluAlaIleHisSerGlnThrGluValaHisLeuLysThrGlyIleValaPro 80
Db 324 TTGGAGAAAGCAATACATTTCCAACTGAAGTTTATTTAAACAGGTGGATTTCTCT 383
Qy 81 ProAlaAlaGlnGlnLeuHisaArgIuAsnIleGlnArgIleValaGlnGluAlaLeuSer 100
Db 384 CCAAGCACTCAACAGCTTCCACAGAGAAATATTCAGAAATATACAGAAAGCTTTTCT 443
Qy 101 AlaSerGlyValaSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeu 120
Db 444 GCCAGTGAAGTCTCTCAAGTACCTCAGCAATTCACATCAATCAAAACAGAGCTT 503
Qy 121 AlaLeuSerLeuGlyValaGlyLeuSerPheSerLeuGluValaGlyGlnLeuLysLys 140
Db 504 GCTTTAAGCCTGGAGTGGCTTATCATTTAGCTTACAGTGTAGAGCACTTAAAAAG 563
Qy 141 ProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLys 160
Db 564 CCAATTCATTCCTTATCATATGAGAGCTCATGCACTTACATTTAGTTGACCAATAA 623
Qy 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180
Db 624 GTAGAAATTCCTTTTATGATCTTTTGAATTCGAGAGTCAGTCTGTGGCATTTAGT 683
Qy 181 GlnGlyValaSerAspPheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet 200
Db 684 CAAGGAGTTTCAGATTTTCTGCTTCTTGGAAGTCTTTGACATAGCACAGGAGCATG 743
Qy 201 LeuAspLysValaAlaArgArgLeuSerLeuIleLysHisProGluCysSerThrMetSer 220
Db 744 CTGACAGAGTGCAGAGAACTTTCTTTATATAACATCCAGAGTCTCCACCATAGT 803
Qy 221 GlyIyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPheHisAspIleLys 240
Db 804 GGTGGAAAGCCATAGAACATTTGGCCAAACAGAAATGATTTCAATTTGACATCAAA 863
Qy 241 ProProLeuHisaIleLysAsnCysAspPheSerPheThrGlyLeuGlnHisValaThr 260
Db 864 CTTCCCTTGATCATGCTTAAATAATGATTTTCTTTTCTGACATTCACACAGTTACT 923
Qy 261 AspLysIleIleMetLysLysGlyLysGluGlnGlyIleGluLysGlyGlnIleLeuSer 280
Db 924 GATTAATAATTAATTAAGAAAGAAAGAAAGAGATTTAGAGAGGGGCAAAATCTGTCT 983
Qy 281 SerAlaAlaAspIleAlaAlaThrValaGlnHisThrMetAlaCysHisLeuValaLysArg 300
Db 984 TCAGCAGACACATGCTGTCACAGTACACACACACATGAGATGATCTCTGTGAAGA 1043
Qy 301 ThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuProGlnAsnAsnAlaVal 320
Db 1044 ACACATCGGCTATTTCTTTTGTAAAGCAGAGACTGTTACTCAAAATTAAGCAGTA 1103
Qy 321 LeuValAlaSerGlyValaAlaSerAsnPhenIleArgArgAlaLeuGluIleLeu 340

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2005, 17:34:34 ; Search time 120 Seconds
(without alignments)
1334.323 Million cell updates/sec

Title: US-10-649-273-2

Perfect score: 414
Sequence: 1 MLITKTAGVFPKSKRRKY.....DISKEVGASIKVPQKMEI 414

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 2105692 seqs, 386760381 residues

Word size: 274

Total number of hits satisfying chosen parameters: 3

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: A_Geneseq_16Dec04:*

- 1: Geneseqp1980s:*
- 2: Geneseqp1980s:*
- 3: Geneseqp2000s:*
- 4: Geneseqp2001s:*
- 5: Geneseqp2002s:*
- 6: Geneseqp2003as:*
- 7: Geneseqp2003bs:*
- 8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	414	100.0	414	5 ABG96478	Abg96478 Novel hum
2	414	100.0	414	5 ABB05481	Abb05481 Human O-s
3	414	100.0	414	6 ABJ26654	Abj26654 Human pro

ALIGNMENTS

RESULT 1

ABG96478 standard; protein; 414 AA.

XX ABG96478;

XX 11-DEC-2002 (first entry)

XX Novel human metalloprotease MPI.

XX Metalloprotease: MP-1; immune disorder; glutamate transport; Cancer;
motor neuron disorder; amyotrophic lateral sclerosis; ALS; diabetes;
reproductive disorder; Klinefelter's syndrome; germinal cell aplasia;
genital wart; metabolic disorder; premature puberty; Kallman syndrome;
Cushing's syndrome; neurodegenerative disease; Alzheimer's disease;
Parkinson's disease; Huntington's disease; Tourette syndrome; sepsis;
liver disease; renal disease; immune disorder; rheumatoid arthritis;

KW acquired immunodeficiency syndrome; AIDS; pulmonary disease; pneumonia;
KW emphysema; cystic fibrosis; vascular disorder; inflammatory disorder;
KW neurological disorder.

OS Homo sapiens.

PN WO200272751-A2.

XX 19-SEP-2002.

PF 05-FEB-2002; 2002WO-US003353.

XX 05-FEB-2001; 2001US-0266518P.

PR 10-APR-2001; 2001US-0282814P.

PA (BRIM) BRISTOL-MYERS SQUIBB CO.

PI Chen J, Feder J, Nelson TC, Duclos F, Krystek S;

XX MPI; 2002-723329/78.

DR N-PSDB; ABS76635.

PT New isolated nucleic acid encoding MP-1 protein, useful for preventing,
PT treating, or ameliorating diseases associated with aberrant
PT metalloproteinase activity, e.g. immune, metabolic, inflammatory and
PT neurological disorders.

PS Claim 5; Fig 1A-C; 473bp; English.

XX The invention describes an isolated nucleic acid molecule (I) encoding a
XX metalloprotease (MP-1). (I) is useful for preventing, treating, or
XX ameliorating a medical condition, particularly an immune disorder, an
XX aberrant glutamate transport or motor neuron disorder, such as
XX amyotrophic lateral sclerosis (ALS), its juvenile form or an ALS-like
XX condition. The compositions and methods are also useful for diagnosing,
XX prognosticating, treating, ameliorating and/or treating disorders
XX associated with MP-1 activity, e.g. diabetes, cancer, reproductive
XX disorders (e.g. Klinefelter's syndrome, genital warts, or germinal cell
XX aplasia), metabolic disorders (e.g. premature puberty, Kallman syndrome,
XX or Cushing's syndrome), neurodegenerative diseases (Alzheimer's disease,
XX Parkinson's disease, Huntington's disease or Tourette syndrome), liver
XX and renal diseases and immune disorders (e.g. AIDS, rheumatoid arthritis
XX or sepsis), pulmonary diseases (e.g. pneumonia, emphysema or cystic
XX fibrosis) and vascular, inflammatory and neurological disorders (e.g.
XX Alzheimer's disease or Parkinson's disease). This is the amino acid
XX sequence of a metalloprotease MPI protein

XX SQ Sequence 414 AA;

Query Match 100.0%; Score 414; DB 5; Length 414;

Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Gaps 0;

Matches 414; Conservative 0; Index 0; Gaps 0;

QY 1 MLITKTAGVFPKSKRRKYEPFRSFPFGTLFLHKIVLGIEISCDTAAAVDRTGNV 60

DB 1 MLITKTAGVFPKSKRRKYEPFRSFPFGTLFLHKIVLGIEISCDTAAAVDRTGNV 60

QY 61 LGBAHSQTEVHLKTGIVPPAAQQLHRENIQRIQVBPASASGVSPDSALATTTIKGL 120

DB 61 LGBAHSQTEVHLKTGIVPPAAQQLHRENIQRIQVBPASASGVSPDSALATTTIKGL 120

QY 121 ALSIGVGLSPSLQVLQKKPPIPIHMEAHALTITLTVNVEPPLVLISGHCILAV 180

DB 121 ALSIGVGLSPSLQVLQKKPPIPIHMEAHALTITLTVNVEPPLVLISGHCILAV 180

QY 181 QGVSPDLGKSLDIPGMDLVNARLSIKRPECSTSGGAIEHTAKQGRFPFDIK 240

DB 181 QGVSPDLGKSLDIPGMDLVNARLSIKRPECSTSGGAIEHTAKQGRFPFDIK 240

QY 241 PPLHAKNCDPSTFGYQHTYTDKIMKEKEBEGIEKQQLISSADIAATVQHTMACHLVKR 300

DB 241 PPLHAKNCDPSTFGYQHTYTDKIMKEKEBEGIEKQQLISSADIAATVQHTMACHLVKR 300

QY 301 THRALLFCORDDLLPONNAVIVASGVASNFYIRRALBITNATCTLLCPPRLCTDNG 360
DB 301 THRALLFCORDDLLPONNAVIVASGVASNFYIRRALBITNATCTLLCPPRLCTDNG 360
QY 361 IMIANGIERLRAGIGLIHIDIEGIRYEPKCPGLVDISKVEGASIKVPOLKMEI 414
DB 361 IMIANGIERLRAGIGLIHIDIEGIRYEPKCPGLVDISKVEGASIKVPOLKMEI 414

RESULT 2

ABR05481
ID ABR05481 standard; protein; 414 AA.

AC ABR05481;

DT 19-APR-2002 (first entry)

DE Human O-sialoglycoproteinase-like protein SEQ ID NO:2.

KM Human, O-sialoglycoproteinase-like protein; OSGPLP; enzyme.

OS Homo sapiens.

PN CN1318550-A.

PD 24-OCT-2001.

PF 19-APR-2000; 2000CN-00106834.

PR 19-APR-2000; 2000CN-00106834.

PA (SHAN-) SHANGHAI SHENGYUAN GENE DEV CO LTD.

PI Mao Y, Xie Y;

DR WPI: 2002-115090/16.

DR N-PSDB; ABA93268.

PT O-sialoglycoproteinase-like protein and encoding polynucleotide, useful for diagnosing, preventing and treating related diseases.

PS Claim 1; Page 30-31 (Disclosure); 38pp; Chinese.

XX The present sequence represents human O-sialoglycoproteinase-like protein (OSGPLP). The present invention also describes: (1) the preparation of the OSGPLP protein; (2) applying the OSGPLP protein in diagnosis; (3) the prevention and/or treatment of related diseases; (4) utilizing the OSGPLP protein in screening its agonist, excitomotor and inhibitor and preparing an antibody against the OSGPLP protein; and (5) the use of the OSGPLP polynucleotide sequences, proteins, agonists, excitomotors, inhibitors and antibodies in treating diseases related to the abnormal OSGPLP gene and in preparing the medicine composite for the treatment

XX Sequence 414 AA;

QY Query Match 100.0%; Score 414; DB 5; Length 414;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLILTKTAGVFFPKSRKRYEEFLRSFNPFGTLFLKIVIGIETSCDDPTAAVVDENGV 60

DB 1 MLILTKTAGVFFPKSRKRYEEFLRSFNPFGTLFLKIVIGIETSCDDPTAAVVDENGV 60

QY 61 LGEALHSQTEVHLKTGIVPPAAQOLAHRENIQRIVOEALASGVSPSDSAIATTIKPGI 120

DB 61 LGEALHSQTEVHLKTGIVPPAAQOLAHRENIQRIVOEALASGVSPSDSAIATTIKPGI 120

QY 121 ALSLGVGLSFSIQLVQOLKKPFIPIHMEAHALTIRLTKVPEPVLVLLISGGHCLIALV 180

DB 121 ALSLGVGLSFSIQLVQOLKKPFIPIHMEAHALTIRLTKVPEPVLVLLISGGHCLIALV 180

QY 181 QGVSPFLILGKSLDIAPGMDLVNARLSLTKPEGSTSGGKALIEHLAQGNRFPHDIK 240

DB 181 QGVSPFLILGKSLDIAPGMDLVNARLSLTKPEGSTSGGKALIEHLAQGNRFPHDIK 240
QY 241 PPLAHAKNCDPSFTGLQHTYTDKIMKKEKEGIEKQIISADIAATVQHTMACHLYKR 300
DB 241 PPLAHAKNCDPSFTGLQHTYTDKIMKKEKEGIEKQIISADIAATVQHTMACHLYKR 300
QY 301 THRALLFCORDDLLPONNAVIVASGVASNFYIRRALBITNATCTLLCPPRLCTDNG 360
DB 301 THRALLFCORDDLLPONNAVIVASGVASNFYIRRALBITNATCTLLCPPRLCTDNG 360
QY 361 IMIANGIERLRAGIGLIHIDIEGIRYEPKCPGLVDISKVEGASIKVPOLKMEI 414
DB 361 IMIANGIERLRAGIGLIHIDIEGIRYEPKCPGLVDISKVEGASIKVPOLKMEI 414

RESULT 3

ABJ2654
ID ABJ2654 standard; protein; 414 AA.

AC ABJ2654;

DT 01-MAY-2003 (first entry)

DE Human protein modification + maintenance molecule protein SEQ ID NO 8.

KM Cytostatic; antiarteriosclerotic; anti-HIV; antiallergic; anticonvulsant;
cerebroprotective; antiparkinsonian; nootropic; antiinflammatory;
antitumor; hepatotropic; gynaecological; antibacterial; vitucide;
protozoacide; antiparasitic; cell proliferative disease; PMOD;
protein modification and maintenance molecule; immunogenic fragment;
cancer; autoimmune; inflammatory disease; neurological disorder;
gastrointestinal; developmental; vesicle trafficking disorder; infection;
protein-protein interaction; drug-target interaction;
gene expression profile; human.

OS Homo sapiens.

PN WO2003000844-A2.

PD 03-JAN-2003.

PF 18-JUN-2002; 2002WO-US019360.

PR 22-JUN-2001; 2001US-0300508P.

PR 06-JUL-2001; 2001US-0303445P.

PR 13-JUL-2001; 2001US-0305405P.

PR 09-AUG-2001; 2001US-0311442P.

PR 24-AUG-2001; 2001US-0314821P.

PR 29-AUG-2001; 2001US-0315992P.

PR 03-MAY-2002; 2002US-0378205P.

XX (INCY-) INCYTE GENOMICS INC.

PA Gandhi AR, Kabir AE, Swarnakar A, Hafalia AJA, Tran B, Duggan BM;

PI Warren BA, Ison CH, Honchell CD, Nguyen DB, Lu DM, Lee EA, Yue H;

PI Forsythe TJ, Barroso I, Ramkumar J, Griffin JA, Li JX, Yang J;

PI Thangaveju K, Glezen KJ, Ding L, Baughn MR, Borowsky MT, Yao MG;

PI Walla NK, Mason PM, Gurrajan R, Lee S, Becha SD, Lee SY, Tran UK;

PI Elliott VS, Luo W, Sprague WW, Tang YT, Lu Y, Zebartadian Y;

XX WPI: 2003-184039/18.
DR N-PSDB; ABT23207.

XX New isolated human PMOD polypeptide and polynucleotide, useful for
PT diagnosing, treating and preventing diseases or conditions associated
PT with the aberrant PMOD expression, e.g. cancer, AIDS, atherosclerosis and
PT infections.

PS Claim 63; Page 182-183; 225pp; English.

XX The invention relates to an isolated polypeptide comprising: any of 28
CC sequences of 48-1256 amino acids; a natural amino acid sequence at least
CC 90% identical to the 28 amino acid sequences, 94% identical to a sequence

CC of 703 or 267 amino acids, 96% identical to a sequence of 414 amino
CC acids, or 97% identical to a sequence of 242 amino acids, all given in
CC the specification; or a biologically active or immunogenic fragment of
CC the isolated polypeptide. The polypeptides and polynucleotides are useful
CC in diagnosing, treating and preventing diseases or conditions associated
CC with the decreased expression of protein modification and maintenance
CC molecules (PMOD), such as cell proliferative diseases (e.g. cancer,
CC atherosclerosis), autoimmune/inflammatory diseases (e.g. AIDS,
CC allergies), neurological disorders (e.g. stroke, Parkinson's disease,
CC epilepsy), gastrointestinal (e.g. ulcer, cirrhosis), reproductive (e.g.
CC endometriosis), developmental, vesicle trafficking disorders, and
CC infections (e.g. bacterial, viral, parasitic, protozoal). These are also
CC useful in assessing the effects of exogenous compounds on the expression
CC of nucleic acid and amino acid sequences of PMOD. The PMOD or its
CC fragments are useful in screening compounds for effectiveness as agonist
CC or antagonist of the polypeptides, or in altering the expression of the
CC target polynucleotide and compounds that specifically bind to or modulate
CC the activity of the polypeptide. The microarray is useful in monitoring
CC or measuring protein-protein interactions, drug-target interactions, and
CC gene expression profiles. This sequence represents a human PMOD protein
CC of the invention
XX

SQ Sequence 414 AA;

Query Match 100.0%; Score 414; DB 6; Length 414;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLITKTGAVFPFKSKRYEFLRSFNHFGTLPLHKIVLGIETSCDDTAAAVVDETGV 60
|||
Db 1 MLITKTGAVFPFKSKRYEFLRSFNHFGTLPLHKIVLGIETSCDDTAAAVVDETGV 60

QY 61 LGEAHSOTEVHAKTGKGIIVPPAAQOLHRENIQIVOEALASGVSPSDISATITKPGI 120
|||
Db 61 LGEAHSOTEVHAKTGKGIIVPPAAQOLHRENIQIVOEALASGVSPSDISATITKPGI 120

QY 121 ALSIGVLSFSLQVLGVOLKKEPIPIHMEBAHALTIRLTNKEFPPLVLLISGHCILALV 180
|||
Db 121 ALSIGVLSFSLQVLGVOLKKEPIPIHMEBAHALTIRLTNKEFPPLVLLISGHCILALV 180

QY 181 QGVSDPFLILGKSLDIAPGMDLKVARRLSLIGHPECTWSGKAIEHLAKQNRFFEDIK 240
|||
Db 181 QGVSDPFLILGKSLDIAPGMDLKVARRLSLIGHPECTWSGKAIEHLAKQNRFFEDIK 240

QY 241 PRLHAKKCDSEFSGLOHVTDKTIIMKKEBEGIEKGQIISAADIAATVQHTMACHLVNR 300
|||
Db 241 PRLHAKKCDSEFSGLOHVTDKTIIMKKEBEGIEKGQIISAADIAATVQHTMACHLVNR 300

QY 301 THRAILFCQKRDILPQNNAVLVASGVASNFYIRRALBITLNATOCITLCPPLRCTDNG 360
|||
Db 301 THRAILFCQKRDILPQNNAVLVASGVASNFYIRRALBITLNATOCITLCPPLRCTDNG 360

QY 361 IMIAMNGIERLAGIILHDIIEGIRYEPKPLGVDISKEVGEASIKVPOLKMEI 414
|||
Db 361 IMIAMNGIERLAGIILHDIIEGIRYEPKPLGVDISKEVGEASIKVPOLKMEI 414

Search completed: November 11, 2005, 03:18:08
Job time : 120 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 11, 2005, 02:14:24 ; Search time 38 Seconds
(without alignments)
1048.256 Million cell updates/sec

Title: US-10-649-273-2

Perfect score: 414
Sequence: 1 MLIIITKTAGVFFPKSKRKRY.....DISKEVGASIKVPOLKWEI 414

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283416 seqs, 96216763 residues

Word size : 274

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR_79: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	ID	Description
------------	-------------	----	-------------

No matches found

Search completed: November 11, 2005, 03:22:58
Job time : 38 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2005, 22:37:04 ; Search time 120 Seconds
(without alignments)
1766.673 Million cell updates/sec

Title: US-10-649-273-2

Perfect score: 414
Sequence: 1 MLILTTAGVFPKPSKRRKY.....DISKEVGASIKVPLQKMEI 414

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1612378 seqs, 512079187 residues

Word size : 274

Total number of hits satisfying chosen parameters: 1

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : UniProt_03:*

1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	414	100.0	414	2 Q96EV9	Q96ev9 homo sapien

ALIGNMENTS

RESULT 1

ID	Q96EV9	PRELIMINARY:	PRT:	414 AA.
AC	Q96EV9			
DT	01-DEC-2001	(TREMBlrel. 19, Created)		
DT	01-DEC-2001	(TREMBlrel. 19, Last sequence update)		
DT	01-MAR-2004	(TREMBlrel. 26, Last annotation update)		
DE	OSGEPL protein.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.			
OX	NCBI_TaxID=9606;			
OX	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Muscle;			
RC	MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;			
RA	Strausberg R.L., Feingold B.A., Grouse L.H., Derge J.G.,			
RA	Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,			
RA	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,			
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,			
RA	Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,			
RA	Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,			
RA	Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,			
RA	Raha S.S., Lequellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,			
RA	Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,			

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.R., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Buttefield Y.S.,
RA Krzywinski M.I., Skalka U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

RP SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RA Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC011904; AA011904.1; -.
DR MEROPS; M22.004; -.
DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000905; Peptidase M22.
DR InterPro; IPR009180; Pept_M22_Osialgl.
DR Pfam; PF00814; Peptidase_M22; 1.
DR PIRSF; PIRSF004537; Osialglc_pptds; 1.
DR PRINTS; PR00789; OSIALOPTASE.
DR ProDom; PD002367; Peptidase_M22; 1.
DR TRIGRAMS; TRIGR00329; gcpd; 1.
SQ SEQUENCE 414 AA; 45122 MW; A536B33F5C6B8BD CRC64;

Query Match 100.0%; Score 414; DB 2; Length 414;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MLILTTAGVFPKPSKRRKYEFPLRSFPHGTLFLHKIVLGIETSCDDTAAVVDGTGNV	60
DB	1	MLILTTAGVFPKPSKRRKYEFPLRSFPHGTLFLHKIVLGIETSCDDTAAVVDGTGNV	60
QY	61	LGSAIHSGTEVHLKGTGIIVPPAAQQLHRENIQRIQVDSASGVSPDSALATTKPGL	120
DB	61	LGSAIHSGTEVHLKGTGIIVPPAAQQLHRENIQRIQVDSASGVSPDSALATTKPGL	120
QY	121	ALSLGVGLSFLQVLQQLKPFPIHMEAHATITRLTNKVEPPVLILISGCHLALV	180
DB	121	ALSLGVGLSFLQVLQQLKPFPIHMEAHATITRLTNKVEPPVLILISGCHLALV	180
QY	181	QGVSDPFLLGKSLIDIAFGDMLDKVARRLSLIKHPECSGSGKAIIEHLAQGRFPHDIK	240
DB	181	QGVSDPFLLGKSLIDIAFGDMLDKVARRLSLIKHPECSGSGKAIIEHLAQGRFPHDIK	240
QY	241	PLIHAANKDFSTGQHTYTDKIIMKKEKEBEGIEKQIILSSADIAATVQHTMACHLYKR	300
DB	241	PLIHAANKDFSTGQHTYTDKIIMKKEKEBEGIEKQIILSSADIAATVQHTMACHLYKR	300
QY	301	THRAILFCORDDLPONNAVVASGVASNYIRALAEITLNTAQTCTLCPPLCTDNG	360
DB	301	THRAILFCORDDLPONNAVVASGVASNYIRALAEITLNTAQTCTLCPPLCTDNG	360
QY	361	IMIANWGIERLRAGLILHIDIEGIRYEPKCPGLVDISKEVGASIKVPLQKMEI	414
DB	361	IMIANWGIERLRAGLILHIDIEGIRYEPKCPGLVDISKEVGASIKVPLQKMEI	414

Search completed: November 11, 2005, 03:22:09
Job time : 120 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 11, 2005, 03:16:49 ; Search time 43 Seconds
(without alignments)
718.714 Million cell updates/sec

Title: US-10-649-273-2

Perfect score: 414
Sequence: 1 MLILTKTAGVFFKPSKRVKRY.....DISKEVEASIKVQLKMEI 414

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 513545 seqs, 74649064 residues

Word size : 274

Total number of hits satisfying chosen parameters: 1

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: listing first 45 summaries

Database : Issued_Patents_AA:*

- 1: /cgn2_6/pdata/1/1aa/5A_COMB.pep:*
- 2: /cgn2_6/pdata/1/1aa/5B_COMB.pep:*
- 3: /cgn2_6/pdata/1/1aa/6A_COMB.pep:*
- 4: /cgn2_6/pdata/1/1aa/6B_COMB.pep:*
- 5: /cgn2_6/pdata/1/1aa/PCTUS_COMB.pep:*
- 6: /cgn2_6/pdata/1/1aa/backfill1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	414	100.0	414	4	US-10-067-443-2 Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-10-067-443-2
; Sequence 2, Application US/10067443
; Patent No. 6642041
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Homo sapiens

US-10-067-443-2

Query Match 100.0%; Score 414; DB 4; Length 414;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MLILTKTAGVFFKPSKRVKRYEFLRSFNFHPTGLFLHKIVLIGISCDPTAAAVDETSNV	60
DB	1	MLILTKTAGVFFKPSKRVKRYEFLRSFNFHPTGLFLHKIVLIGISCDPTAAAVDETSNV	60
QY	61	LGSAIHSOTVHLKTKGIVPPAAOOLRENIQRIVOEPLSASGVSPSLAIAATTIKPGL	120
DB	61	LGSAIHSOTVHLKTKGIVPPAAOOLRENIQRIVOEPLSASGVSPSLAIAATTIKPGL	120
QY	121	ALSLGVLSFSLQVLVQLKKPFIPIHMEAHALITRLTNKVEPFLVLLISGHCIALAV	180
DB	121	ALSLGVLSFSLQVLVQLKKPFIPIHMEAHALITRLTNKVEPFLVLLISGHCIALAV	180
QY	181	QGVSDPFLILGKSLDIAPGMDLKVARRLSLIKHECSTMSGKAI EHLAKQGNRPHFDIK	240
DB	181	QGVSDPFLILGKSLDIAPGMDLKVARRLSLIKHECSTMSGKAI EHLAKQGNRPHFDIK	240
QY	241	PPLHAKNCPESFTGLQHTVDKTIIMKKEKEGIERKQILSSAADIAATVQHTMACHLVKR	300
DB	241	PPLHAKNCPESFTGLQHTVDKTIIMKKEKEGIERKQILSSAADIAATVQHTMACHLVKR	300
QY	301	THRAILFCXORDLLPQNNAVLVASGVASNFYIRRALEILTNTAQTCLLCPPLCTDNG	360
DB	301	THRAILFCXORDLLPQNNAVLVASGVASNFYIRRALEILTNTAQTCLLCPPLCTDNG	360
QY	361	IMIAMNGIERLRAGLGIHDIIGIRYEPKCPGLGVDISKEVEASIKVQLKMEI 414	
DB	361	IMIAMNGIERLRAGLGIHDIIGIRYEPKCPGLGVDISKEVEASIKVQLKMEI 414	

Search completed: November 11, 2005, 03:23:52
Job time : 43 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 11, 2005, 03:17:09 ; Search time 113 Seconds

(without alignments)
1532.934 Million cell updates/sec

Title: US-10-649-273-2

Perfect score: 414
Sequence: 1 MLILTKTAGVFFPKSRKRVY.....DISKEVGASIKVPLKMEI 414

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1867879 seqs, 418409474 residues

Word size : 274

Total number of hits satisfying chosen parameters: 4

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Published Applications AA:*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTOS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US10E_PUBCOMB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
19: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep.*
20: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
21: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
22: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	414	100.0	414	US-10-067-443-2	Sequence 2, Appli
2	414	100.0	414	US-10-649-273-2	Sequence 2, Appli
3	414	100.0	414	US-10-651-722-2	Sequence 2, Appli
4	414	100.0	414	US-10-480-988-8	Sequence 8, Appli

ALIGNMENTS

RESULT 1
US-10-067-443-2

Sequence 2, Application US/10067443
Publication No. US20030082782A1
GENERAL INFORMATION:
APPLICANT: Bristol-Myers Squibb Company
TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED
TITLE OF INVENTION: SPINAL CORD, MP-1
FILE REFERENCE: D0073 NP
CURRENT APPLICATION NUMBER: US/10/067,443
CURRENT FILING DATE: 2002-02-05
PRIOR APPLICATION NUMBER: US 60/266,518
PRIOR FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: US 60/282,814
PRIOR FILING DATE: 2001-04-10
NUMBER OF SEQ ID NOS: 71
SOFTWARE: PatentIn version 3.0
SEQ ID NO 2
LENGTH: 414
TYPE: PRT
ORGANISM: Homo sapiens
US-10-067-443-2

Query Match 100.0%; Score 414; DB 14; Length 414;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLILTKTAGVFFPKSRKRVYFLRSFNFHFGTLFLHKIVGIETSCDDTAAYVDENGV 60
DB 1 MLILTKTAGVFFPKSRKRVYFLRSFNFHFGTLFLHKIVGIETSCDDTAAYVDENGV 60

QY 61 LGEAHSQTEVHLKGTGIVPAAQOLHRENIQRIQVQALSASGVSPDLSAIAATTIKGL 120
DB 61 LGEAHSQTEVHLKGTGIVPAAQOLHRENIQRIQVQALSASGVSPDLSAIAATTIKGL 120

QY 121 ALSIGLSTSLQVQOLKRPPIPHMEHAATLRTNKVBPPLVLLISGHCILAV 180
DB 121 ALSIGLSTSLQVQOLKRPPIPHMEHAATLRTNKVBPPLVLLISGHCILAV 180

QY 181 QGVSDFLKGSJLDIAPGMDLVARLSLTKHPECSGSGKAIBHLAQGRFPHDIK 240
DB 181 QGVSDFLKGSJLDIAPGMDLVARLSLTKHPECSGSGKAIBHLAQGRFPHDIK 240

QY 241 PPLHAKNCDPFTGLQHTVTDKIMKKEKEGIEKGQILSSADIAATVQHTWACHLYKR 300
DB 241 PPLHAKNCDPFTGLQHTVTDKIMKKEKEGIEKGQILSSADIAATVQHTWACHLYKR 300

QY 301 THRALLFCQORDLLPQNNAVLVASGVASNFYIRRALEIITNAOTCTLLCPPRLCTDNG 360
DB 301 THRALLFCQORDLLPQNNAVLVASGVASNFYIRRALEIITNAOTCTLLCPPRLCTDNG 360

QY 361 IMIANGGIRLRAGLGLIHDIIEGIRYBPCKPLGVDISKEVGASIKVPLKMEI 414
DB 361 IMIANGGIRLRAGLGLIHDIIEGIRYBPCKPLGVDISKEVGASIKVPLKMEI 414

RESULT 2
US-10-649-273-2

Sequence 2, Application US/10649273
Publication No. US20040043407A1
GENERAL INFORMATION:
APPLICANT: Bristol-Myers Squibb Company
TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1
FILE REFERENCE: D0073 CNT
CURRENT APPLICATION NUMBER: US/10/649,273
CURRENT FILING DATE: 2003-08-27
PRIOR APPLICATION NUMBER: US 60/266,518
PRIOR FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: US 10/067,443
PRIOR FILING DATE: 2002-02-05
PRIOR APPLICATION NUMBER: US 60/282,814
PRIOR FILING DATE: 2001-04-10
NUMBER OF SEQ ID NOS: 71
SOFTWARE: PatentIn version 3.2
SEQ ID NO 2

```

; LENGTH: 414
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-649-273-2

Query Match          100.0%; Score 414; DB 15; Length 414;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLILTKTAGVFPFKSRKRYEFLRSFNFHPTGLFLHKIVLGIFSCDDTAAAVDETGNV 60
   |||
DB 1 MLILTKTAGVFPFKSRKRYEFLRSFNFHPTGLFLHKIVLGIFSCDDTAAAVDETGNV 60

QY 61 LGEAHSQTEVHLKKGIVPPAAQOLHRENIGRIYOEALASASGVSPDSALATTTKPL 120
   |||
DB 61 LGEAHSQTEVHLKKGIVPPAAQOLHRENIGRIYOEALASASGVSPDSALATTTKPL 120

QY 121 ALSIGVGSFSLQVLGOLKKPPIPIHMEBAHALTTRLTKVFPFLVLLISGHCILALV 180
   |||
DB 121 ALSIGVGSFSLQVLGOLKKPPIPIHMEBAHALTTRLTKVFPFLVLLISGHCILALV 180

QY 181 QGVSDPFLLGKSLDIAPGMDLKVARRSLIKHPECSTWSGKALIEHLAKQGNRFHFDIK 240
   |||
DB 181 QGVSDPFLLGKSLDIAPGMDLKVARRSLIKHPECSTWSGKALIEHLAKQGNRFHFDIK 240

QY 241 PPLHAKNCDFSFTGLQHTYTDKIIMKEKEBEGIEKQIISADIAATVOHTMACHLVVR 300
   |||
DB 241 PPLHAKNCDFSFTGLQHTYTDKIIMKEKEBEGIEKQIISADIAATVOHTMACHLVVR 300

QY 301 THRALLFCORDLPLPNNNAVLVASGVASNFIYRRLALETITNATQCTLLCPPRLCTDNG 360
   |||
DB 301 THRALLFCORDLPLPNNNAVLVASGVASNFIYRRLALETITNATQCTLLCPPRLCTDNG 360

QY 361 IMIANGIERLRAGLGIHLHDIGIRYEPKPLGVDISKVEGASIRVPOLKMEI 414
   |||
DB 361 IMIANGIERLRAGLGIHLHDIGIRYEPKPLGVDISKVEGASIRVPOLKMEI 414

RESULT 3
US-10-651-722-2
; Sequence 2, Application US/10651722
; Publication No. US20040048302A1
; GENERAL INFORMATION:
; APPLICANT: Bistrol -Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1
; FILE REFERENCE: D0073 DIV
; CURRENT APPLICATION NUMBER: US/10/651,722
; CURRENT FILING DATE: 2003-08-29
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-651-722-2

Query Match          100.0%; Score 414; DB 15; Length 414;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLILTKTAGVFPFKSRKRYEFLRSFNFHPTGLFLHKIVLGIFSCDDTAAAVDETGNV 60
   |||
DB 1 MLILTKTAGVFPFKSRKRYEFLRSFNFHPTGLFLHKIVLGIFSCDDTAAAVDETGNV 60

QY 61 LGEAHSQTEVHLKKGIVPPAAQOLHRENIGRIYOEALASASGVSPDSALATTTKPL 120
   |||
DB 61 LGEAHSQTEVHLKKGIVPPAAQOLHRENIGRIYOEALASASGVSPDSALATTTKPL 120
```

```

QY 121 ALSIGVGSFSLQVLGOLKKPPIPIHMEBAHALTTRLTKVFPFLVLLISGHCILALV 180
   |||
DB 121 ALSIGVGSFSLQVLGOLKKPPIPIHMEBAHALTTRLTKVFPFLVLLISGHCILALV 180

QY 181 QGVSDPFLLGKSLDIAPGMDLKVARRSLIKHPECSTWSGKALIEHLAKQGNRFHFDIK 240
   |||
DB 181 QGVSDPFLLGKSLDIAPGMDLKVARRSLIKHPECSTWSGKALIEHLAKQGNRFHFDIK 240

QY 241 PPLHAKNCDFSFTGLQHTYTDKIIMKEKEBEGIEKQIISADIAATVOHTMACHLVVR 300
   |||
DB 241 PPLHAKNCDFSFTGLQHTYTDKIIMKEKEBEGIEKQIISADIAATVOHTMACHLVVR 300

QY 301 THRALLFCORDLPLPNNNAVLVASGVASNFIYRRLALETITNATQCTLLCPPRLCTDNG 360
   |||
DB 301 THRALLFCORDLPLPNNNAVLVASGVASNFIYRRLALETITNATQCTLLCPPRLCTDNG 360

QY 361 IMIANGIERLRAGLGIHLHDIGIRYEPKPLGVDISKVEGASIRVPOLKMEI 414
   |||
DB 361 IMIANGIERLRAGLGIHLHDIGIRYEPKPLGVDISKVEGASIRVPOLKMEI 414

RESULT 4
US-10-480-988-8
; Sequence 8, Application US/10480988
; Publication No. US20050069877A1
; GENERAL INFORMATION:
; APPLICANT: GANDHI, Ameena R.; KABIE, Amy E.;
; APPLICANT: SWARNAKAR, Anita; HARALIA, April J.A.;
; APPLICANT: TRAN, Bao; DUGAN, Brendan M.;
; APPLICANT: WARREN, Bridget A.; ISON, Craig H.;
; APPLICANT: HONCHELL, Cynthia D.; NGUYEN, Danielle B.;
; APPLICANT: LU, Dyung Aina M.; LEE, Ernestine A.;
; APPLICANT: YUE, Henry; FORSYTHE, Ian J.;
; APPLICANT: BARROSO, Ines; RAMKUMAR, Jayalaxmi;
; APPLICANT: GRIFFIN, Jennifer A.; LI, Joana X.;
; APPLICANT: YANG, Junming; THANGAVELU, Kavitha;
; APPLICANT: GIETZEN, Kimberly J.; DING, Li;
; APPLICANT: BAUGHN, Mariah R.; BOROMSKY, Mark L.;
; APPLICANT: YAO, Monique G.; CHAWLA, Natinder K.;
; APPLICANT: MASON, Patricia M.; GURURAJAN, Rajagopal;
; APPLICANT: LEE, Sally; BECHA, Shanya D.;
; APPLICANT: LEE, Soo Yeun; TRAN, Uyen K.;
; APPLICANT: ELLIOTT, Vicki S.; LHO, Wen;
; APPLICANT: SPRAGUE, William W.; TANG, Y. Tom;
; APPLICANT: LU, Yan; ZEBARUDJIAN, Yeganeh
; TITLE OF INVENTION: PROTEIN MODIFICATION AND MAINTENANCE MOLECULES
; FILE REFERENCE: PF-1040 USN
; CURRENT APPLICATION NUMBER: US/10/480,988
; CURRENT FILING DATE: 2003-12-16
; PRIOR APPLICATION NUMBER: PCT/US02/19360
; PRIOR FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: US 60/300,508
; PRIOR FILING DATE: 2001-06-22
; PRIOR APPLICATION NUMBER: US 60/303,445
; PRIOR FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US 60/305,405
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/311,442
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: US 60/314,821
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/315,992
; PRIOR FILING DATE: 2001-08-29
; PRIOR APPLICATION NUMBER: US 60/378,205
; PRIOR FILING DATE: 2002-05-03
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: PERL Program
; SEQ ID NO 8
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
```

; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No: 7632424CD1
US-10-480-988-8

Query Match 100.0%; Score 414; DB 17; Length 414;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MLILTKTAGVFPFKSRKRYEFLRSFNFPGLFLHKIVLGIETSCDDTAADVDETNV	60
Db	1	MLILTKTAGVFPFKSRKRYEFLRSFNFPGLFLHKIVLGIETSCDDTAADVDETNV	60
Qy	61	LGEAHSQTEVHLKTGGIVPPAQQLRHNIRIVQEAALSAGVSPSDLSAIAITTIKPGI	120
Db	61	LGEAHSQTEVHLKTGGIVPPAQQLRHNIRIVQEAALSAGVSPSDLSAIAITTIKPGI	120
Qy	121	ALSLGVGLSFSLOLVGOLKKPPIPIHMEAHALTRLTKVFPPLVLLISGHCILALV	180
Db	121	ALSLGVGLSFSLOLVGOLKKPPIPIHMEAHALTRLTKVFPPLVLLISGHCILALV	180
Qy	181	QGVSDFLILGKSLDIPGMDLKVARRLSLIKHPECSTWSSGKAIEHLAKOGNRFHDIK	240
Db	181	QGVSDFLILGKSLDIPGMDLKVARRLSLIKHPECSTWSSGKAIEHLAKOGNRFHDIK	240
Qy	241	PPLHAKNCDFSFTGLQHVTDKIIMKKEKEGIEKGQILSSADIAATVQHTMACHLYR	300
Db	241	PPLHAKNCDFSFTGLQHVTDKIIMKKEKEGIEKGQILSSADIAATVQHTMACHLYR	300
Qy	301	THRAILFCQRDLIPQNNNAVVASGVASNFYIRRALIELTNAOTCTLLCPPRLCTDNG	360
Db	301	THRAILFCQRDLIPQNNNAVVASGVASNFYIRRALIELTNAOTCTLLCPPRLCTDNG	360
Qy	361	IMIAMNGIERLRAGILGILHDIEGIRYEPKCPGVDISKEVGEASIKVPOLKMEI	414
Db	361	IMIAMNGIERLRAGILGILHDIEGIRYEPKCPGVDISKEVGEASIKVPOLKMEI	414

Search completed: November 11, 2005, 03:25:57
Job time : 113 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 11, 2005, 03:23:55 ; Search time 295 Seconds
(without alignments)
2296.334 Million cell updates/sec

Title: US-10-649-273-2

Perfect score: 414
Sequence: 1 MLILTKTAGVFEPKSKRRKY.....DISKEVGASIRVQLKMEI 414

Scoring table:

	OLIGO	Xgapop 60.0	Xgapext 60.0
	Xgapop 60.0	Xgapext 60.0	
	Fgapop 6.0	Fgapext 7.0	
	Delop 6.0	Delext 7.0	

Searched: 1202784 seqs, 818138359 residues

Word size: 274

Total number of hits satisfying chosen parameters: 2

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2_1/USPRO_spool_p/US10649273/runat_02112005_091641_16903/app_query.fasta.1.583
-DB=Issued_Patents_NA -QFWT=fascap -SUFFIX=01gpn.rn1 -MINMATCH=0.1 -LOOPC1=0
-LOOPEXT=0 -UNITS=bites -STAR=1 -END=1 -MATRIX=01gpn -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=quality -THR MIN=274 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptio -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10649273@cgn 1.1.105@runat_02112005_091641_16903 -NCRU=6 -ICPU=3
-NO_MMAP -LARGEOUTERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:*

1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	414	100.0	2197	4	US-10-067-443-1 Sequence 1 Appli
2	340	82.1	1416	4	US-09-774-528-177 Sequence 177, App

ALIGNMENTS

RESULT 1
US-10-067-443-1
; Sequence 1, Application US/10067443
; Patent No. 6642041
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company

;; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED
;; FILE OF INVENTION: SPINAL CORD, MP-1
;; FILE REFERENCE: D0073 NP
;; CURRENT APPLICATION NUMBER: US/10/067,443
;; CURRENT FILING DATE: 2002-02-05
;; PRIOR APPLICATION NUMBER: US 60/266,518
;; PRIOR FILING DATE: 2001-02-05
;; PRIOR APPLICATION NUMBER: US 60/282,814
;; PRIOR FILING DATE: 2001-04-10
;; NUMBER OF SEQ ID NOS: 71
;; SOFTWARE: PatentIn version 3.0
;; SEQ ID NO: 1
;; LENGTH: 2197
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (231)..(1472)
US-10-067-443-1

Alignment Scores:

Pred. No.:	Score:	Percent Similarity:	Best Local Similarity:	Query Match:	DB:
0	414.00	100.00%	100.00%	100.00%	4
	Matches:	414	Conservative:	0	
	Mismatches:	0	Indels:	0	
	Gaps:	0			

US-10-649-273-2 (1-414) x US-10-067-443-1 (1-2197)

QY	1	MetleuileuThryStralaglyValPhepelysProserLysArglyValTyr	20
DB	231	ATGCTATCTTGACTAGCTAGCTGAGAGGTTTTTTTAAACATCAAAAGAAAGTTAT	290
QY	21	GluPheleuArgSerPheasnPheHISProGlyThreuleuPheulHslySileValleu	40
DB	291	GAATTTTAAAGATTATTTATTTCTCTGAAACATATTTCTTCAATAAATAGTATTG	350
QY	41	GlylleGluThserCysasPepThrAlaAlaValValPepGluThrglyValVal	60
DB	351	GGAATTTGAACTAGTGTGATGATACGACGCTGTGTGTGATGAACTGAAATGTG	410
QY	61	LeuGlyGluAlaIleHISserGlnThrglyValHISleuysThrglyGlyIleValPro	80
DB	411	TTGGAGAACCATATCATCTCCAACTGAACTTATTAACACAGGTGGATTGTTCT	470
QY	81	ProAlaIleGlnGlnleuHISargGluasniIleGlnargIleValGlnGlnAlaSer	100
DB	471	CCAGCAGCTCAACAGCTTCAAGAGAAATATTCACAGATAGTACAAAGAGCTTTCT	530
QY	101	AlaSerGlyValSerProserAspLeuSerAlaIleAlaThrIlelyPProGlyLeu	120
DB	531	GCAGAGGAGTCTCTCAAGTACCTCTCAAGCAATTCACATCAATTAACAGAGCTT	590
QY	121	AlaLeuSerleuGlyValGlyLeuSerPheSerleuGlnleuValGlyGlnleuLys	140
DB	591	GCTTAAGCTGGAGGTGGCTTATCATTTAGCTTACAGCTGGTGAACGTTAAAG	650
QY	141	ProPheIleProIleHISmetGluAlaHISAlaIleuThrIleArgleuThrAsnLys	160
DB	651	CCATTCAATCCCATTCATATATGAGGGCTCAATCTTATTTAGTTAGTCAATTA	710
QY	161	ValGluPheProPheleuValleuLysIleSerGlyGlyHISCySleuLeuAlaIleVal	180
DB	711	GTAAGATTTCTTTTATTTAGTTCTTTGATTTCTGAGGCTCAGCTGTGTGGATTAGT	770
QY	181	GlnGlyValSerAspPheleuLeuGlyLysSerleuAspIleAlaProGlyAspMet	200
DB	771	CAAGGATTTCAAGATTTTCTTTCTTGAAGCTTTTGGACATAGGACCAAGGATAG	930
QY	201	LeuAspLysValAlaArgArgleuSerleuIlelyHISProGluCySerThrMetSer	220
DB	831	CTTGACAAAGTGGCAAGAGACTTTCTTTAATTAACATTCAGAGTCTCCACCATG	890

QY 221 GYGLYLSAIAIEGLUHIISLEUALALYSGLINGLYASARGPHEHISPHESPILELYS 240
 DB 891 GGTGGAAAGCCATAGAACATTTGGCCAAACAGAAATAGATTTCAATTTGACATCAAA 950
 QY 241 PROPLEUHIISHIALYASNCYSAPPHESERPHETHGLYLEUGLNIISVALIHR 260
 DB 951 CCTCCCTTGACATCATGCTAAATAATTTGATTTTCTTTTACTGGACTTCAACCGTTACT 1010
 QY 261 ASPLYSILEIEMETLYLSGLULYSGTUGLUGLYILEGLULYSGLYGLNILELEUSER 280
 DB 1011 GATATAATATATATGAAAAAGAAAAAGAGAAAGATTGAGAAAGGGGAAATCTGTCT 1070
 QY 281 SERIAAIAAPPIIEALALATHRVALGINHISTHMEALCYSHISLEUVALLYSARG 300
 DB 1071 TCAGAGACGACATGTCTGCACAGACACACACATGACATGTCATCTTGTAAGAA 1130
 QY 301 THRHSAGALAIIELEUPHESYSGINARGASPLEULEUPROGLNANANALAVAI 320
 DB 1131 ACACATCGGCGTATCTGTTTGTGAAGCAGAGACTTTTACTCAAAATATGACAGTA 1190
 QY 321 LEUVALIASERGLYGLYVALIASERANPHEIYRIEARGALALEUGLIULEU 340
 DB 1191 CTGGTTGATCTGGTGTGTCCAGTAAGTATCTATATCCGACAGACTCTGAAATTTTA 1250
 QY 341 THRANALATHRINCYSERHLEULEUCYSPPROPPOARGLEUCYSTRASPAENGLY 360
 DB 1251 ACAAACGACACACATGCACTTGTGTGTCTCTCCACGACTATGACCTGTAATATGCG 1310
 QY 361 ILEMETIEALATPRASNGLYILEGLUARGLEUARGALGLYLEUGLYILEUHIISAP 380
 DB 1311 ATTATGATTCATGATGAAATGTATTTGAAACATACGCTGGCTGGGCAATTTTACATGAC 1370
 QY 381 IIEGLUGLYIEARGTYGTIGUPROLYSCYSPROLEUGLYVALASPILESERLYSGIULAI 400
 DB 1371 ATAGAAGGATCGCTATGACCAAAATGCTCTTGAGATGACATATCAAAAGATT 1430
 QY 401 GLYGLUALASERILELYSVALPROGLNLEULEUYSMETGLIULE 414
 DB 1431 GGAGAGCTTCATTAAGTACCAATTAATAATGAGATA 1472
 RESULT 2
 US-09-774-528-177
 ; Sequence 177, Application US/09774528
 ; Patent No. 6743619
 ; GENERAL INFORMATION:
 ; APPLICANT: Tang, Y. Tom
 ; APPLICANT: Zhou, Ping
 ; APPLICANT: Goodrich, Kyle
 ; APPLICANT: Liu, Chenghua
 ; APPLICANT: Asundi, Vinod
 ; APPLICANT: Ren, Feiyan
 ; APPLICANT: Zhang, Jie
 ; APPLICANT: Zhao, Qing A.
 ; APPLICANT: Yang, Yonghong
 ; APPLICANT: Xue, Aidong J.
 ; APPLICANT: Wehrman, Tom
 ; APPLICANT: Wang, Jian-Rui
 ; APPLICANT: Wang, Dunrui
 ; APPLICANT: Drmanac, Radoje T.
 ; TITLE OF INVENTION: No. 6743619el Nucleic Acids and
 ; FILE REFERENCE: 802
 ; CURRENT APPLICATION NUMBER: US/09/774, 528
 ; NUMBER OF SEQ ID NOS: 441
 ; SOFTWARE: pc_fl_genes Version 2.0
 ; SEQ ID NO 177
 ; LENGTH: 1416
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: CDS

; LOCATION: (205)..(1305)
 US-09-774-528-177
 Alignment Scores:
 Pred. No.: 0
 Score: 340.00
 Percent Similarity: 100.00%
 Best Local Similarity: 100.00%
 Query Match: 82.13%
 DB: 4
 Gaps: 0
 US-10-649-273-2 (1-414) x US-09-774-528-177 (1-1416)
 QY 75 THRGLYGLYILEVALPROPROAIAIAGLNGLEUHIISARGLUBENILEGINARILE 94
 DB 283 ACAGGTGGATTTCTTCTCCAGCAGCTCAACACCTTCACAGAAATATTCACAGATA 342
 QY 95 VALGLINGLUALLEUSERALASERGLYVALSERPROSERAPLESERAIIEALATHR 114
 DB 343 GTACAAAGAGCTTTCTGCGACAGTGTCTCCAAAGTACCTTCAGCAATTTGCAACT 402
 QY 115 THRILEYSPROGLYLEUALALEUSERLEUGLYVALGLYLEUSERPHESERLEUGLNU 134
 DB 403 ACCATTAACACAGACTTGTCTTAAGCTGGAGTGGCTTATCATTTACCTTACAGCTG 462
 QY 135 VALGLYGLNLEUYSYSPROPHIELEPROILEHISHIMETGLUALAHSIALALEUTHR 154
 DB 463 GTAGACAGATTAAAGCCATTCATTCATTCATATGAGGCTCATGCACTTACT 522
 QY 155 IIEARGLEUTHRANLYSVALGIUPHETPROPHLEUVALLEUUILESERGLYGLYHIS 174
 DB 523 ATTAGGTGACCATATAAGTAGAATTTCTTTTATGTTCTTTGATTTCTGGAGTGCAC 582
 QY 175 CYSEULEUALALEUVALINGLYVALSERASPHLEULEULEUGLYLYSERLEUASP 194
 DB 583 TGTCTGTGTCATTAGTCAAGAGATTTCAGATTTTCGTTCTTGGAAAGCTTTTGAC 642
 QY 195 IIEALAPROGLYASPHETLEUASPLYSVALIARGRLEUSERLEUILEYSHISPRO 214
 DB 643 ATAGCACACAGTACATGCTTGACAGGTGCAAGAACTTTTATTAATTAACATCA 702
 QY 215 GLUCYSERTHMETSERGLYGLYLSAIAIEGLUHIISLEUALALYSGLINGLYASARG 234
 DB 703 GAGTGTCCACACATGATGCTGGGAAAGCCATGAACTTTGGCCAAACAGAAATGAA 762
 QY 235 PHEHISPHESAPPILEYSPROPLEUHIISHIALYASNCYSAPPHESERPHETHR 254
 DB 763 TTTCAATTTGACATCAAACTCCCTTGACATCATGCTAAAAATGTGATTTTCTTTACT 822
 QY 255 GLYLEUGLNIISVALIHRASPLYILEIEMETLYLSGLULYSGIUGLUGLYILEGLU 274
 DB 823 GGACTTCACACAGTTCATGATTAATAATATGAAAAAGAAAAAGAAAGGATTTGAG 882
 QY 275 LYSGLYGLNILEUSERSERALASERGLYVALIIEALATHRVALGINHISTHMEALIA 294
 DB 883 AAGGGGAAATCTGTTTTCAGCAGCAGCATTTCTCCACAGACACACATGCA 942
 QY 295 CYSHISLEUVALLYSARGTHRSARGALAIIELEUPHESYSGINARGASPLEULEU 314
 DB 943 TGTATCTTGGAAGAAACACACATCGGCTATTTCTGTTTGTAGACAGAGACTTGTTA 1002
 QY 315 PROGLNANANALAVALLLEUVALIASERGLYGLYVALIASERANPHEIYRIEARG 334
 DB 1003 CCTCAAAATATATGACAGTGTGATCTGGTGTGTGCGAAGTAACCTTTATATCCGC 1062
 QY 335 ARGALALEUGLIULEUTHRANALATHRGLNCYSERHLEULEUCYSPPROPPOARG 354
 DB 1063 AGAGCTTCGAAATTTTAAACAAACGACACAGTGCATTTGTGTGCTCTCCCGA 1122
 QY 355 LEUCYSTRASPAENGLYILEMETIEALATPRASNGLYILEGLUARGLEUARGALGLY 374
 DB 1123 CTATGACATGATATGACATTTATGATTCATGATGATGATTTGAAAGACTAGTGTGC 1182


```

Qy      375 LeuGlyIleLeuHisAspIleGluGlyIleArgTyrGluProIleGlySerProLeuGlyVal 394
          |||||
Db      1183 TTGGGCATTTTACATGACATGAGGCAATCCGCTATGAACCAAAATGTCCTCTTGGAGTA 1242
          |||||
Qy      395 AspIleSerIleGluValGlyGluIleAspIleValProGlnLeuIleMetGluIle 414
          |||||
Db      1243 GACATATCAAAAGAAAGTTGGAGAGCTTCCATTAAGGTACACACATTTAAATAATGAGATA 1302
          |||||

```

Search completed: November 11, 2005, 07:21:00
 Job time : 298 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 11, 2005, 03:18:15 ; Search time 887 Seconds
(without alignments)
2762.991 Million cell updates/sec

Title: US-10-649-273-2

Perfect score: 414
Sequence: 1 MLILTKTAGVFFKPSKRYK.....DISKEVGASIVPQLKMEI 414

Scoring table:

OLIGO	Xgapop 60.0	Xgapext 60.0
Ygapop 60.0	Ygapext 60.0	
Fgapop 6.0	Fgapext 7.0	
Delop 6.0	Delext 7.0	

Searched: 4390206 seqs, 2959870667 residues

Word size: 274

Total number of hits satisfying chosen parameters: 4

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n model -DEV=xlp
-Q=/cgn2_1/USFTO_spool_p/US10649273/runat_02112005_031640_16863/app_query.fasta_1.583
-DB=N_Geneseq_16Dec04 -QFMT=fastap -SUFFIX=oligp2n.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=quality -THR MIN=274 -ALIGN=15 -MODS=LOCAL
-OUTFMT=plco -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USRR=US10649273 @CGN 1.1 708 @runat_02112005_031640_16863 -NCPU=6 -ICPU=3
-NO_MMAR -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOC
-DEV_TIMEOUT=10 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

N_Geneseq_16Dec04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	414	100.0	2058	6	ABA93268 Human O-8
2	414	100.0	2197	6	AB876635 DNA encod
3	414	100.0	2572	8	ABT23207 Human pro
4	340	82.1	1416	8	ABX70950 Novel hum

ALIGNMENTS

RESULT 1

ABA93268
ID ABA93268 standard; cDNA; 2058 BP.

AC ABA93268;

DT 19-APR-2002 (first entry)

DE Human O-sialoglycoproteinase-like protein encoding cDNA SEQ ID NO:1.

KW Human; O-sialoglycoproteinase-like protein; OSGPUP; enzyme; gene; ss.

OS Homo sapiens.

FT Key location/Qualifiers

FT CDS 110..1354

FT /*tag= a

PN CN1318550-A.

PD 24-OCT-2001.

PP 19-APR-2000; 2000CN-00106834.

PR 19-APR-2000; 2000CN-00106834.

XX (SHAN-) SHANGHAI SHENGYUAN GENE DEV CO LTD.

XX Mao Y, Xie Y;

XX WPI, 2002-115090/16.

XX DR P-PSDB; ABB05481.

PT O-sialoglycoproteinase-like protein and encoding polynucleotide, useful for diagnosing, preventing and treating related diseases.

PS Claim 5; Page 29-30 (Disclosure); 38pp; Chinese.

XX The present sequence encodes human O-sialoglycoproteinase-like protein (OSGPUP). The present invention also describes: (1) the preparation of the OSGPUP protein; (2) applying the OSGPUP protein in diagnosis; (3) the prevention and/or treatment of related diseases; (4) utilizing the OSGPUP

CC protein in screening its agonist, excitomotor and inhibitor and preparing an antibody against the OSGPUP protein; and (5) the use of the OSGPUP

CC polynucleotide sequences, proteins, agonists, excitomoters, inhibitors

CC and antibodies in treating diseases related to the abnormal OSGPUP gene

XX and in preparing the medicine composite for the treatment

SQ Sequence 2058 BP; 637 A; 400 C; 410 G; 611 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	0	Length:	2058
Score:	414.00	Matches:	414
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-10-649-273-2 (1-414) x ABA93268 (1-2058)

QY	1	MetLeuIleLeuThrIysThraIaGlyValPhePheIysProSerIysArgIysValTyr	20
DB	110	ATGCTAATCTTGAAGCTGACGAGGATTTTAAACCAACCAAGAAAGTTAT	169
QY	21	GlnPheLeuArgSerPheLeuPheIlePrcGlyThrLeuPheLeuIleIysIleValIleu	40
DB	170	GAATTTTAAAGATTATTTTTCATCCGAAACATATTTCTTCATAAATAAGTATG	229
QY	41	GlyIleGluThrSerCysAspPheThrAlaAlaValValAspGluThrGlyAsnVal	60

```

Db      230 GGAATTGAAACAGTGTGATGATACAGACGCTGCTGTGTGATGAAACTGGAAATGTG 289
Qy      61 LeuGIyGluValaIleHisSerGlnThrGluValHisLeuLysThrGlyYlleValPro 80
Db      290 TTGGGAGAGACAAATCATTCCTCCAACTGAAGTTCATTTAAAAACAGTGGAGATTTGTTCT 349
Qy      81 ProIlaaIaGlnGlnLeuHisArgGluAenIIeGlnArgIleValGlnGlnValaLeuSer 100
Db      350 CCAGACGCTCAACAGCTTCACAGAGAAATATTCAGAAAGATGATACAGAGAGCTTTTCT 409
Qy      101 AlaSerGIyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGIyLeu 120
Db      410 GCCAGTGAAGTCTCTCCAAAGTACCTCTCAGCAATTGCACTACCAATAAAACAGAGACTT 469
Qy      121 AlaLeuSerLeuGIyValGIyLeuSerPheSerLeuGlnLeuValGIyGlnLeuLysLys 140
Db      470 GCTTTAAGCTCTGGAGTGGCTTATCATTTAGCTTACAGCTGGTAGAGACGTTAAAAAG 529
Qy      141 ProPheIleProIleHisHisMetGluAaHisAlaLeuThrIleArgLeuThrAsnLys 160
Db      530 CCATTCAATTCATCATCATATGAGAGGCTCAGTCACTATATAGTTGACCAATTA 589
Qy      161 ValGIuPheProPheLeuValLeuLeuIleSerGIyYHisCysLeuLeuAlaLeuVal 180
Db      590 GTAGAAATTCCTTTTGTAGTTCTTTGATTTCTGGAGGTCACTGCTGTTGGCATTTAGTT 649
Qy      181 GlnGIyValSerAspPheLeuLeuLeuGIyLysSerLeuAspIleAlaProGIyAspMet 200
Db      650 CAAGAGATTTCAAGATTTTCTGCTTTGGAAAGTCTTTGGACATACACAGGCTGACATG 709
Qy      201 LeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGluCysSerThrMetSer 220
Db      710 CTGGAAGAAGTGGCAAGAGACTTTCTTATATAAACATCCAGAGTCTCCACCATGAGT 769
Qy      221 GLyGIyLysAlaIleGlnHisLeuAlaLysGlnGIyAsnArgPheHisPheAspIleLys 240
Db      770 GGTGGAGAAAGCCATTAACATTTGGCCAAACAGAAATAGATTTGATTTGACATCMA 829
Qy      241 ProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGIyLeuGlnHisValThr 260
Db      830 CCTCCCTTGTCATCATCTCAAAAATTTGATTTTCTTTTACTGGACTTCAACACGTTACT 889
Qy      261 AspLysIleIleMetLysLysGIyLysGIyGlnGIyIleGlnLysGIyGlnIleLeuSer 280
Db      890 GATAAATATATATGAAAAAGGAAAAAGGAAAGTATGAGAAAGGGGCAAAATCTGCTCT 949
Qy      281 SerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCysHisLeuValLysArg 300
Db      950 TCAGCAGCAGACATTCCTGTCACAGACAGACACACATGTCATCTTTGTGAAAGA 1009
Qy      301 ThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuProGlnAsnAlaVal 320
Db      1010 ACRATCGGGCTATTTCTGTTTGTAAAGCAGAGACTTGTACTCAAAATATATGACGTA 1069
Qy      321 LeuValAlaSerGIyGIyValAlaAspAsnPheThrIleArgArgAlaLeuGlnIleLeu 340
Db      1070 CTGGTTCATCTGGTGGTGTCCGACAGTAATCTTATATCCGACAGACTCTGCGAAATTTTA 1129
Qy      341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeuCysThrAspAsnGIy 360
Db      1130 ACMAAGCAGACACAGTCACTTTGTTGTGTCTCTCCACAGCTATGCACCTGATTAATGGC 1189
Qy      361 IleMetIleAlaIleArgAsnGIyIleGluArgLeuArgAlaGIyLeuGIyIleLeuHisAsp 380
Db      1190 ATTTATATTCATGATGAAATGTTATGAAAGCTACGTCCTGGCTTGGCATTTTATCAGAC 1249
Qy      381 IleGlnGIyIleArgTyrGlnProLysCysProLeuGIyValAspIleSerLysGlnVal 400
Db      1250 ATAGAAAGCATCGGCTATGAAACAAATGTCTCTTGGATGACATATCAAAAAGAGTT 1309
Qy      401 GlyGluAlaSerIleLysValProGlnLeuLysMetGlnIle 414

```

```

Db      1310 GGAGAGCTTCATATAAAGTACCAATTAATAAATGAGATA 1351
RESULT 2
AB576635
ID      AB576635 standard; DNA; 2197 BP.
XX
AC      AB576635;
XX
DT      11-DEC-2002 (first entry)
XX
DE      DNA encoding novel human metalloproteinase MPI.
XX
XX      Metalloproteinase: MP-1; immune disorder; glutamate transport; cancer;
KW      motor neuron disorder; amyotrophic lateral sclerosis; ALS; diabetes;
KW      reproductive disorder; Klinefelter's syndrome; germinal cell aplasia;
KW      genital wart; metabolic disorder; premature puberty; Kallman syndrome;
KW      Cushing's syndrome; neurodegenerative disease; Alzheimer's disease;
KW      Parkinson's disease; Huntington's disease; Tourette syndrome; sepsis;
KW      liver disease; renal disease; immune disorder; rheumatoid arthritis;
KW      acquired immunodeficiency syndrome; AIDS; pulmonary disease; pneumonia;
KW      emphysema; cystic fibrosis; vascular disorder; inflammatory disorder;
KW      neurological disorder; gene; ds.
XX
XX      Homo sapiens.
XX
XX      PN      W0200272751-A2.
XX
XX      PD      19-SEP-2002.
XX
XX      PF      05-FEB-2002; 2002MO-US003353.
XX
XX      PR      05-FEB-2001; 2001US-0266518P.
XX      PR      10-APR-2001; 2001US-0282814P.
XX
XX      PA      (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX
XX      PI      Chen J, Feder J, Nelson TC, Duclos F, Krystek S;
XX      WPI; 2002-723329/78.
XX      DR      P-PSDB; ABG96478.
XX
XX      PT      New isolated nucleic acid encoding MP-1 protein, useful for preventing,
PT      treating, or ameliorating diseases associated with aberrant
PT      metalloproteinase activity, e.g. immune, metabolic, inflammatory and
PT      neurological disorders.
XX
XX      PS      Claim 1; Fig 1A-C; 473bp; English.
XX
XX      CC      The invention describes an isolated nucleic acid molecule (I) encoding a
CC      metalloproteinase (MP-1). (I) is useful for preventing, treating, or
CC      ameliorating a medical condition, particularly an immune disorder, an
CC      aberrant glutamate transport or motor neuron disorder, such as
CC      amyotrophic lateral sclerosis (ALS), its juvenile form or an ALS-like
CC      condition. The compositions and methods are also useful for diagnosing,
CC      prognosticating, treating, ameliorating and/or treating disorders
CC      associated with MP-1 activity, e.g. diabetes, cancer, reproductive
CC      disorders (e.g. Klinefelter's syndrome, genital wart, or germinal cell
CC      aplasia), metabolic disorders (e.g. premature puberty, Kallman syndrome,
CC      or Cushing's syndrome), neurodegenerative diseases (Alzheimer's disease,
CC      Parkinson's disease, Huntington's disease or Tourette syndrome), liver
CC      and renal diseases and immune disorders (e.g. AIDS, rheumatoid arthritis
CC      or sepsis), pulmonary diseases (e.g. pneumonia, emphysema or cystic
CC      fibrosis) and vascular, inflammatory and neurological disorders (e.g.
CC      Alzheimer's disease or Parkinson's disease). This sequence represents a
CC      metalloproteinase MPI polynucleotide
XX
XX      SO      Sequence 2197 BP; 681 A; 441 C; 439 G; 636 T; 0 U; 0 Other;

```

Alignment Scores: 0
 Pred. No.: 2197
 Score: 414.00
 Percent Similarity: 100.00%
 Best Local Similarity: 100.00%
 Length: 2197
 Matches: 414
 Conservative: 0
 Mismatches: 0

```
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-10-649-273-2 (1-414) x ABS76635 (1-2197)

Qy 1 MetLeuIleuThrThyThraIaGlyValaPhePheIysProSerIysArgIysValTyr 20
Db 231 ATGCTAACTCTTGACCTAGACAGCTGACAGAGATTTTAAACCAACCAAAAGAAAGTTTAT 290
Qy 21 GluPheLeuAArgSerPheAsnPhenHisProGlyThrLeuPheLeuHisIleValLeu 40
Db 291 GAATTTTAAAGAGTTTATTTTATTTCTCTGGAACCACTATTTCTTATATAATAGATTTG 350
Qy 41 GlyIleGluThrSerCysAspAspThrAlaAlaValValAspGluThrGlyAsnVal 60
Db 351 GGAATTGAACCTAGTTGTGATGATACAGACAGCTGCTGTGGATGAATGAACCTGGAATGTG 410
Qy 61 LeuGlyGluAlaIleHisSerGlnThrGluValaHisIleuIysThrGlyIleValPro 80
Db 411 TTGGAGAGAGCAATACATTTCCCAACCTGAAGTTCAATTAAACACAGTGGATTTGTTCT 470
Qy 81 ProAlaIaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100
Db 471 CCAGCAGCTCAACAGCTTCAAGAGAAATATTCACGAAATAGTACAGAGAGCTTTTCT 530
Qy 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleIysProGlyLeu 120
Db 531 GCCAGTGGAGTCTCTCCAAAGTGAAGCTTCCAGCAATTTGCAATACCAATAAACAGAGACTT 590
Qy 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuIysIys 140
Db 591 GCTTTAAAGCTGGAGTGGCTTATCATTTACTTACAGCTGATGAGACAGTTAAAGAAAG 650
Qy 141 ProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeuThrAsnIys 160
Db 651 CCATTCATTTCCCATTCATCATATGAGAGGCTCATGACCTTACATTAAGTTGACCAATAAA 710
Qy 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180
Db 711 GTAGAAATTTCTTTTAACTTTTAACTTTTGAATTTCTGAGAGTCACTGTCTGTGGCATTAAGT 770
Qy 181 GlnGlyValSerAspPheLeuLeuLeuGlyIysSerLeuAspIleAlaProGlyAspMet 200
Db 771 CAAGAGATTTCAAGATTTCTGCTCTTGGAAAGCTTTTGAACATAGCACAGGTGACATG 830
Qy 201 LeuAspIysValAlaArgArgLeuSerIleIysHisIleProGluCysSerThrMetSer 220
Db 831 CTTACCAAGGTGGCAAGAAAGACTTTCTTAAATMAAACATCCAGAGTGTCCACCATGAGT 890
Qy 221 GlyGlyIysAlaIleGlnHisIleuAlaIysGlnGlyAsnArgPheHisPheAspIleIys 240
Db 891 GGTGGAGAAAGCATAGAACATTTGGCCAAACAGAAATAGATTTTACATTCACATCAAA 950
Qy 241 ProProLeuHisHisAlaIysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260
Db 951 CCTCCCTTGATCATAGCTAAAAATGATATTTTCTTTTACAGCACTTCAACAGCTTACT 1010
Qy 261 AspIysIleIleMetIysIysGluIysGluGlyIleGlnIysGlyGlnIleLeuSer 280
Db 1011 GATAAATTAATTAATGAAGAAAGGAAAGGAGGATATTTGAGAGAGGCAATTCCTGTCT 1070
Qy 281 SerAlaIaAspIleAlaAlaThrValGlnHisIleThrMetAlaCysHisIleuValIysArg 300
Db 1071 TCAGAGAGAGCAATTTGCTGCAACAGTACAGCACAAATGAGCAATGTCATTTTGAAAGAA 1130
Qy 301 ThrHisArgAlaIleLeuPheCysIysGlnArgAspLeuLeuProGlnIleAsnAlaVal 320
Db 1131 ACACATCGGCTATTTCTTTTGTAAAGCAGAGACTTCTTACCTCAAAATATATGACAGTA 1190
Qy 321 LeuValAlaSerGlyIysValAlaSerAsnPhenIleArgArgAlaLeuGluIleLeu 340
Db 1191 CTGGTTGATCTGGGTGTGTGCAAGTAATCTTATATCCGACAGCTCTGCAAAATTTTA 1250
```

```
Qy 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProAlaArgLeuCysThrAspAsnGly 360
Db 1251 ACAACGCCAACACAGTGCACCTTTGTGTGTCTCTCTCCAGACTATACATGATAATGGC 1310
Qy 361 IleMetIleAlaThrPheAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380
Db 1311 ATTAATGATTCATGAGATGAGATGATTAAGAAAGCTAGTCTGCTGGCATTTTACATGAC 1370
Qy 381 IleGluGlyIleArgIleGluProIysCysProLeuGlyValAspIleSerIysGluVal 400
Db 1371 ATAGAAGCATTCGGCTATGAAACCAAAATGTCCTTTGAGATGACATATCAAAAGAGTT 1430
Qy 401 GlyGluAlaSerIleIysValProGlnLeuIysMetGluIle 414
Db 1431 GGAGAAAGCTTCCATTAAGATACACAAATTAATAATGAGAGATA 1472

RESULT 3
ABT23207
ID ABT23207 standard; DNA; 2572 BP.
XX
AC ABT23207;
XX
DT 01-MAY-2003 (first entry)
XX
DE Human protein modification + maintenance molecule DNA SEQ ID No 36.
XX
KW Cytostatic; antiarteriosclerotic; anti-HIV; antiallergic; anticonvulsant;
KW cerebroprotective; antiparkinsonian; neurotropic; antiinflammatory;
KW antitumor; hepatotropic; gynaecological; antibacterial; virucide;
KW protozoacide; antiparasitic; cell proliferative disease; PMOD;
KW protein modification and maintenance molecule; immunogenic fragment;
KW cancer; autoimmune; inflammatory disease; neurological disorder;
KW gastrointestinal; developmental; vesicle trafficking disorder; infection;
KW protein-protein interaction; drug-target interaction;
KW gene expression profile; human; gene; ds.
XX
OS Homo sapiens.
XX
PN MO2003000844-A2.
XX
PD 03-JAN-2003.
XX
PF 18-JUN-2002; 2002MO-US019360.
XX
PR 22-JUN-2001; 2001US-0300508P.
PR 06-JUL-2001; 2001US-0303445P.
PR 13-JUL-2001; 2001US-0305405P.
PR 09-AUG-2001; 2001US-0311442P.
PR 24-AUG-2001; 2001US-03114821P.
PR 29-AUG-2001; 2001US-0315992P.
PR 03-MAY-2002; 2002US-0378205P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Gandhi AR, Kable AE, Swarnakar A, Hafalia AJA, Tran B, Duggan BM;
PI Warren BA, Ison CH, Honchell CD, Nguyen DB, Lu DAM, Lee BA, Yue H;
PI Forsthe IJ, Barroso I, Rankumar J, Griffitt JA, Li JX, Yang J, McG;
PI Thangaveilu K, Gietzen KJ, Ding L, Baughn MK, Borowsky ML, Yao S;
PI Walla NK, Mason PM, Gunturajan R, Lee S, Becha SD, Lee SY, Tran UK;
PI Billicot VS, Luo W, Sprague WW, Tang YT, Lu Y, Zebajadian Y;
XX
DR WPI: 2003-184039/18.
DR P-PSDB; AB026654.
XX
PT New isolated human PMOD polypeptide and polynucleotide, useful for
PT diagnosing, treating and preventing diseases or conditions associated
PT with the aberrant PMOD expression, e.g. cancer, AIDS, atherosclerosis and
PT infections.
XX
PS Claim 91; Page 211; 225pp; English.
XX
CC The invention relates to an isolated polypeptide comprising: any of 28
sequences of 48-1256 amino acids; a natural amino acid sequence at least
```

CC 90% identical to the 28 amino acid sequences, 94% identical to a sequence
CC of 703 or 267 amino acids, 96% identical to a sequence of 414 amino
CC acids, or 97% identical to a sequence of 242 amino acids, all given in
CC the specification; or a biologically active or immunogenic fragment of
CC the isolated polypeptide. The polypeptides and polynucleotides are useful
CC in diagnosing, treating and preventing diseases or conditions associated
CC with the decreased expression of protein modification and maintenance
CC molecules (PMD), such as cell proliferative diseases (e.g. cancer,
CC atherosclerosis), autoimmune/inflammatory diseases (e.g. AIDS,
CC allergies), neurological disorders (e.g. stroke, Parkinson's disease,
CC epilepsy), gastrointestinal (e.g. ulcer, cirrhosis), reproductive (e.g.
CC endometriosis), developmental, vesicle trafficking disorders, and
CC infections (e.g. bacterial, viral, parasitic, protozoal). These are also
CC useful in assessing the effects of exogenous compounds on the expression
CC of nucleic acid and amino acid sequences of PMD. The PMD or its
CC fragments are useful in screening compounds for effectiveness as agonist
CC or antagonist of the polypeptides, or in altering the expression of the
CC target polynucleotide and compounds that specifically bind to or modulate
CC the activity of the polypeptide. The microarray is useful in monitoring
CC or measuring protein-protein interactions, drug-target interactions, and
CC gene expression profiles. This polynucleotide sequence represents the DNA
CC encoding a human PMD protein of the invention

XX Sequence 2572 BP; 780 A; 489 C; 509 G; 794 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 0 Length: 2572
Score: 414.00 Matches: 414
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 8 Gaps: 0

US-10-649-273-2 (1-414) x ABT23207 (1-2572)

QY 1 MetLeuIleLeuThrLysThrAlaGlyValPhePheLeuProSerLysArgLysValTyr 20
DB 144 ATGCTAATCTTGACATAAGCTCAGAGAGTTTAAAAACATCAAAAAGAAAGTTTAT 203
QY 21 GluPheLeuArgSerPheAsnPhenHisProGlyThrLeuPheLeuHisLysIleValLeu 40
DB 204 GAATTTTAAAGAAAGTTTAAATTTTCTCTGGAACACTATTTCTTATTAATAAATGATTTG 263
QY 41 GlyIleGluThrSerCysAspAspThrAlaAlaValValAspGluThrGlyAsnVal 60
DB 264 GGAATTTGAAACTGATGTGATATACAGACGCTGCTGTGATTAACCTGAAATGTG 323
QY 61 LeuGlyGluAlaIleHisSerGlnThrGluValHisLeuLysThrGlyIleValPro 80
DB 324 TTGGAGAAAGCAATACATTCCTCAAGCTGAAGTTCAATTAATAAACAAGTGGATTTCTCT 383
QY 81 ProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100
DB 384 CCAGACAGCTCAACAGCTTCACAGAGAAATATTCAAGAAATAGTACAAAGAGCTCTTTCT 443
QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeu 120
DB 444 GCCAGTGGAGTCTCTCCCAAGTACCTCTCAGCAAAATGCAACTACATAAAACAGAGCTT 503
QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLysLys 140
DB 504 GCTTTAAGCTCGGAGAGTGGCTTATCATTTAGCTTCAAGCTGTAGAGACAGTTAAAAAG 563
QY 141 ProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLys 160
DB 564 CCATTCAATCCCATTCATCATATGAGGCTCAGCACTTATGTTAGTTGACCAATATAA 623
QY 161 ValGluPheProPheLeuValLeuLeuLysSerGlyGlyHisCysLeuLeuAlaLeuVal 180
DB 624 GTAGAAATTTCTTTTATTTTATTTCTTTTGAATTTCTGAGGAGCACTGTGTTGGAATTA 683
QY 181 GlnGlyValSerAspPheLeuLeuGlnGlyLysSerLeuAspIleAlaProGlyAspMet 200

DB 684 CAAAGAGTTTCAGATTTTCTGCTTTCGAAAAGCTTTTGGACATACAGAGTGACATG 743
QY 201 LeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGluCysSerThrMetSer 220
DB 744 CTGGACAAAGTGGCAAGAAAGACTTTCTTTATTAATAAATCAGAGTGTCTCACCATAGT 803
QY 221 GlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLys 240
DB 804 GGTGGAAAGCCATATAGAACTTTGGCCAAACAAAGAAATAGATTTCTTTTGAATCAATCAA 863
QY 241 ProProLeuHisHisAlaLysAsnGlyAspPheSerPheThrGlyLeuGlnHisValThr 260
DB 864 CTTCCCTTGATATATGCTAAAAATTTGATTTTCTTTTATCTGAACTTCAACACGTTACT 923
QY 261 AspLysIleIleMetLysLysGlyLysGlnGluGlyIleGluLysGlyGlnIleLeuSer 280
DB 924 GATTAATAATTAATGAAAAAGAAAAAGAAAGGATTTAGAAAGGGCAAAATCTGTCT 983
QY 281 SerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCysHisLeuValLysArg 300
DB 984 TCAGCAGACAGACATTTGCTCCACAGTACAGACACATATGCGATGTCTTGTGAAGAAGA 1043
QY 301 ThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuProGlnAsnAlaVal 320
DB 1044 ACACATCGGGCTATTTCTGTTTGTAAAGCAGACAGACTTGTACCTCAAAATTAATGACGTA 1103
QY 321 LeuValAlaSerGlyGlyValAlaSerAspPheTyrIleArgArgAlaLeuGluIleLeu 340
DB 1104 CTGTGTGATCTGTGGTGTGTGCGAAGTAACTTTATATTCGCGAGAGCTCTGGAATTTTA 1163
QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeuCysThrAspAsnGly 360
DB 1164 ACAAGCAGAACACAGGCACTTTGTTGTGTCTCTCCAGACTATGCACTGATTAATGCGC 1223
QY 361 IleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380
DB 1224 ATATATGATTCATGATGATGATGATGAAAGCTAGCTGCTGTGGCATTTTATCATGAC 1283
QY 381 IleGluGlyIleArgTyrGluProLysCysProLeuGlyValAspIleSerLysGluVal 400
DB 1284 ATGAAAGGCATCCGCTATGAAACCAAAATGTCTCTTGAAGTATGACATATCAAAAGATT 1343
QY 401 GlyIleAlaSerIleLysValProGlnLeuLysMetGluIle 414
DB 1344 GGAAGAGCTTCATTAATAAGTACCAATTAATAAATGAGATA 1385
RESULT 4
ID ABX70950 standard; cDNA; 1416 BP.
AC ABX70950;
XX
XX 05-MAR-2003 (first entry)
XX
DE Novel human cDNA sequence #175.
XX
XX Human; gene; ss; nervous system disorder; peripheral neuropathy;
KW Huntington's disease; amyotrophic lateral sclerosis; haemophilia;
KW neurodegenerative disease; Parkinson's disease; Alzheimer's disease;
KW autoimmune disease; systemic lupus erythematosus; rheumatoid arthritis;
KW insulin-dependent diabetes mellitus; anaemia; chromocyclopaenia; wound;
KW ulcer; burn; bone disorder; osteoporosis; osteoarthritis; stroke;
KW fibrosis; reperfusion injury; infection; allergic rhinitis; asthma;
KW coagulation disorder; cancer; tumour; inflammatory disease; septic shock;
KW Crohn's disease; anaphylaxis; proliferation; chemotactic;
KW differentiation; stem cell growth factor; haematopoiesis; chemokinetic;
KW haemostatic; antiinflammatory; expressed sequence tag; EST.
OS Homo sapiens.
XX
XX NO200281731-A2.
PN
XX
PD 17-OCT-2002.

XX 29-JAN-2002; 2002WO-US001222.
 XX
 XX 30-JAN-2001; 2001US-00774528. (HYSE-) HXSEQ INC.
 XX (GOOD/) GOODRICH R W.
 XX
 XX Tang TY, Liu C, Zhou P, Asundi V, Zhang J, Zhao Q, Ren F;
 PI Xue A, Yang Y, Wehrman T, Wang J, Wang D, Drmanac R;
 XX WPI; 2003-058563/05.
 XX
 PT Novel polypeptide useful for treating neurodegenerative diseases, myeloid
 PT or lymphoid cell disorders, bone disorders, mechanical and traumatic
 PT disorders, coagulation disorders, and inflammatory diseases.
 XX
 PS Claim 1; Page: 612pp; English.

CC This invention relates to the cDNA sequences encoding an isolated novel
 CC human polypeptide. The protein encoded by the nucleic acid of the
 CC invention is useful for treating central and peripheral nervous system
 CC diseases (e.g. peripheral neuropathy, Huntington's disease, amyotrophic
 CC lateral sclerosis); neurodegenerative diseases (e.g. Parkinson's disease,
 CC Alzheimer's disease); autoimmune disease (e.g. systemic lupus
 CC erythematosus, rheumatoid arthritis, insulin-dependent diabetes mellitus)
 CC myeloid or lymphoid cell disorders (e.g. anaemia and thrombocytopenia)
 CC wounds, ulcers, burns; bone disorders (e.g. osteoporosis,
 CC osteoarthritis); mechanical and traumatic disorders (e.g. stroke, head
 CC trauma); lung or liver fibrosis; reperfusion injury in various tissues;
 CC bacterial, viral or fungal infections; allergic conditions such as
 CC allergic rhinitis, asthma; coagulation disorders (e.g. haemophilia);
 CC cancer and tumours; and inflammatory diseases (e.g. septic shock, Crohn's
 CC disease, anaphylaxis). The protein may be used to inhibit the growth,
 CC infection or function of infectious agents such as bacteria, fungi,
 CC viruses, or to effect bodily characteristics, biorhythms or circadian
 CC cycles of rhythms. The protein may also have
 CC proliferation/differentiation, stem cell growth factor, haematopoiesis
 CC regulation, immune stimulation or suppressing, chemotactic/chemokinetic,
 CC haemostatic and thrombolytic, receptor/ligand, and antiinflammatory
 CC activities. The cDNA sequences of the invention are useful for expressing
 CC recombinant protein for analysis. The present sequence represents a novel
 CC human cDNA sequence of the invention, this sequence is an expressed
 CC sequence tag (EST) and was identified using subtractive hybridisation
 XX
 XX Sequence 1416 BP; 441 A; 280 C; 273 G; 422 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 0 Length: 1416
 Score: 340.00 Matches: 340
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 82.13% Indels: 0
 Gaps: 0

US-10-649-273-2 (1-414) x ABX70950 (1-1416)

QY 75 ThrGlyGlyIleValProProAlaAlaGlnGlnLeuHisArgGlnAsnIleGlnArgIle 94
 DB 283 ACAGGTGGATGTCTCCAGCAGCTCAACAGCTTCAAGAAATATTCACGAATA 342
 QY 95 ValGlnGlnAlaLeuSerAlaSerGlyValSerProSerAspLeuSerAlaIleAlaThr 114
 DB 343 GTACAAAGAGCTCTTTCGACAGTGAAGTCTCTCCAAAGTACCTCTCAGCAATTCACACT 402
 QY 115 ThrIleLysProGlyLeuAlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeu 134
 DB 403 ACCATTAAACAGACAGCTTAAAGCTTGAAGTGGCTTATCATTTAGCTTACAGCTG 462
 QY 135 ValGlyGlnLeuLysLysProPheIleProIleHisIleHisIleMetGlnAlaHisIleLeuThr 154
 DB 463 GTAGGACAGTTAAAGAGCCATTCATTCATTCATCATCATATGAGAGCTCATGCACTTACT 522

QY 155 IleArgLeuThrAspLysValGluPheProPheLeuValLeuLeuIleSerGlyGlyHis 174
 DB 523 ATTAGGTGACCAATAAAGTAAAGATTCCTTTTCTAGTCTTTGATTTCTGAGGTCAC 582
 QY 175 CysLeuLeuAlaLeuValGlnGlnValSerAspPheLeuLeuLeuGlyLysSerLeuAsp 194
 DB 583 TGTCTGTGGCATTAGTTCAGAGAGTTCAGATTTTCTCTTCTTGGAAAGCTTTTGGAC 642
 QY 195 IleAlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisPro 214
 DB 643 ATAGCACACAGTGAATGCTTGAACAGTGGCAAGAAAGACTTTTATATTAACATCCA 702
 QY 215 GluCysSerThrMetSerGlyGlyLysAlaIleMetLysLysGlnLysGlnGlyIleGlu 234
 DB 703 GAGTCTCCACCATGAGTGTGGAAAGCCATAGAACATTTGGCCAAACAGAAATAGA 762
 QY 235 PheHisPheAspIleLysProProLeuHisHisAlaLysAsnGlyAspPheSerPheThr 254
 DB 763 TTTCATTTTGAACATCAAACTCCCTGCAATCATGCTAAATAATGATTTTCTTTTACT 822
 QY 255 GlyLeuGlnHisValThrAspLysIleIleMetLysLysGlnLysGlnGlyIleGlu 274
 DB 823 GAGCTTCAACACGTTACTGATTAATATATATGAAAAAGAAAGAAAGGATTTAG 882
 QY 275 LysGlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAla 294
 DB 883 AAGGGCAATCTGTTCTTACGACGACAGATTCCTCCACAGTACGACACATGCGCA 942
 QY 295 CysHisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeu 314
 DB 943 TGTCACTTGTGAAAAAGAACACATCGGCTATTTCTGTTTGTAAACAGAGACTGTGA 1002
 QY 315 ProGlnAspAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheTyrIleArg 334
 DB 1003 CCTCAAAATATATACAGTACTGTTGCACTGTGTGTGTCACAGTAACTTTATATATCGC 1062
 QY 335 ArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArg 354
 DB 1063 AGAGCTCTGGAATTTTAAACAAACGACACAGTGCCTTTGTTGTCTCTCCACGA 1122
 QY 355 LeuCysThrAspAsnGlyIleMetIleAlaIleArgAsnGlyIleGlnArgLeuArgAlaGly 374
 DB 1123 CTATGACATGATATAGCATTTATGATTCATGGAATGATTTGAAGAATACGTCGTCGC 1182
 QY 375 LeuGlyIleLeuHisAspIleGlnGlyIleArgTyrGlnProLysCysProLeuGlyVal 394
 DB 1183 TTGGGCAATTTTATCATGACATAGAAAGCATCCGCTATGAAACAAATGTCTCTTGGAGTA 1242
 QY 395 AspIleSerLysGlnValGlyGlnAlaSerIleLysValProGlnLeuLysMetGlnIle 414
 DB 1243 GACATATCAAAAGAGTTGGAGAGCTTCCATTAAGTACCAATTAATAATGAGATA 1302

Search completed: November 11, 2005, 03:41:03
 Job time : 895 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 11, 2005, 03:23:05 ; Search time 5350 Seconds
(without alignments)
2945.534 Million cell updates/sec

Title: US-10-649-273-2
Perfect score: 414
Sequence: 1 MLILTKTAGVFFPKSKRKY.....DISKEVGASIVPOLKMEI 414

Scoring table:

OLIGO	
Xgapop 60.0 , Xgapext 60.0	
Ygapop 60.0 , Ygapext 60.0	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 34239544 segs, 19032134700 residues

Word size: 274

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlp
-O=/cgn2.1/USPRO.spool.p/US10649273/runat.02112005.091641.16886/app.query.fasta_1.583
-DB=EST -QFMT=fastap -SUFFIX=oligp2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCCALIGN=200 -THR SCORE=quality -THR MIN=274 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10649273 @CGN 1 1 5180 @runat.02112005.091641.16886 -NCPU=6 -ICPU=3
-NO MMAP -LARGESQLERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOC
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_hic:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description

No matches found					

Search completed: November 11, 2005, 07:15:50
Job time : 5350 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 11, 2005, 03:22:15 ; Search time 7521 Seconds

(without alignments)
2667.260 Million cell updates/sec

Title: US-10-649-273-2

Perfect score: 414
Sequence: 1 MLILTKTAGVFFPKSRKRY.....DISKEVGASIRVPLQKMEI 414

Scoring table: OLIGO

Xgapop 60.0	Xgapext 60.0
Fgapop 60.0	Fgapext 60.0
Delop 6.0	Delext 7.0

Searched: 4708233 seqs, 24227607955 residues

Word size: 274

Total number of hits satisfying chosen parameters: 3

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: listing first 45 summaries

Command line parameters:

```
-MODSL=frame+ p2n model -DEV=xlp
-Q=/cgn2.1/USFPO_spool_p/US10649273/runat_02112005_091640_16872/app_query.fasta_1.583
-DB=genEmbl -QFMT=fastap -SUFFIX=ol1gp2n.rge -MINMATCH=0.1 -LOOPECL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=ol1go -TRANS=human40.cdi
-LIST=45 -DOCCALIGN=200 -THR SCORE=quality -THR MIN=274 -ALIGN=15 -MODE=LOCAL
-OUTMT=plco -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10649273 @CGN 1.1 5600 @runat_02112005_091640_16872 -NCPU=6 -ICPU=3
-NO_MMAP -LRQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOC
-DEV_TIMEOUT=150 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7
```

Database :

```
GenEmbl:*
1: gb_da:*
2: gb_hlg:*
3: gb_in:*
4: gb_cm:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_ets:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	414	100.0	BC011904	Homo sapi
2	414	100.0	AR428803	Sequence
3	340	82.1	AR541929	Sequence

ALIGNMENTS

RESULT 1
LOCUS BC011904
DEFINITION Homo sapiens O-sialoglycoprotein endopeptidase-like 1, mRNA (CDNA clone MGC:20293 IMAGE:4121450), complete cds.
ACCESSION BC011904
VERSION BC011904.2 GI:40225818
KEYWORDS MGC.
SOURCE Homo sapiens (human).
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 1908)
Straussberg, R.L., Peingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shennan, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Bueltow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullaly, S.J., Bosak, S.A., McKwan, P.J., McKernan, K.J., Malek, J.A., Guaratine, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalls, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A., Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

TITLE

12477932 2 (bases 1 to 1908)
Straussberg, R.

JOURNAL

Submitted (30-JUL-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REFERENCE

12477932 2 (bases 1 to 1908)
Straussberg, R.

AUTHORS

Submitted (30-JUL-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

JOURNAL

Submitted (30-JUL-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK

NIH-MGC Project URL: <http://mgc.nci.nih.gov>
On Dec 19, 2003 this sequence version replaced gi:15080281.

COMMENT

Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: <http://www.nisc.nih.gov/>
Contact: nisc_mgc@nih.gov
Ahter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, E., Khong, P., Laric, P., Legaspi, R., Maduro, Q.L., Masello, C., Maskeli, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Stancirpop, S., Thomas, P.J., Touchman, J.W., Tsurgouon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
Series: IRAL Plate: 28 Row: 3 Column: 22.
Location/Qualifiers
1. 1908
/organism="Homo sapiens"
/mol_type="mRNA"

FEATURES

source

Score: 414.00 Matches: 414
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-649-273-2 (1-414) x AR428803 (1-2197)

QY 1 MetLeuIleLeuThrLysThrAlaGlyValPhePheLysProSerLysArgLysValIle 20
Db 231 ATGCTAACTTGCACAGACGCTGAGAGCTTTTAAACCATCAAAAAGAAAGTTAT 290
QY 21 GluPheLeuArgSerPheAsnPhenHisProGlyLysLeuPheLeuHisLysIleValIle 40
Db 291 GAATTTTAAAGAGTTTAAATTTTCACTGGAACCTATTTCTTATAAAATAGATG 350
QY 41 GlyIleGluThrSerCysAspAspThrAlaAlaValAspGluThrGlyAsnVal 60
Db 351 GGAATTGAACTAGTGTGATGATACAGACGCTGCTGTGTGATGAACTGGAAATGTG 410
QY 61 LeuGlyGluAlaIleHisSerGlnThrGlyValHisLeuLysThrGlyIleValPro 80
Db 411 TTGGAGAGAGCAATATCATTTCCAACTGAAGTTCAATTAACAGGTGGATGTTCT 470
QY 81 ProAlaAlaGlnGlnLeuHisArgGlyLeuHisLeuArgIleValGlnGluAlaLeuSer 100
Db 471 CCACAGCTCAACAGCTTCAACAGAGAAATATTTCAACGATATGACAGAGCTTTCT 530
QY 101 AlaSerGlyValSerProSerAspSerLeuSerAlaIleAlaThrThrIleLysProGlyLeu 120
Db 531 GCCAGTGAAGTCTTCCAAAGTACCTCTGAGCAATGCACTACCAATAAACAGAGACTT 590
QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLysLys 140
Db 591 GCTTTAAGCTGGAGAGTGGCTTATCATTTACCTTACAGCTGTGAGACAGTTAAAG 650
QY 141 ProPheIleProIleHisHisMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLys 160
Db 651 CCATTCATTTCCATTCATCATATGAGGCTCATGACATTAATTAAGTTGACCAATAA 710
QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180
Db 711 GTAAATTTCTTTTAACTTTTGAATTCGTGAGGCTGCTGTGTGACATTAAGT 770
QY 181 GlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspGluAlaProGlyAspMet 200
Db 771 CAAGAGCTTCAAGATTTCTGCTTCTTGGAAAGCTTTTGAACATACACAGAGTGACATG 830
QY 201 LeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGluCysSerThrMetSer 220
Db 831 CTTCACAGAGTGGCAAGAAAGCTTTCTTTAATTAACATCCAGAGTGTCCACATGAGT 890
QY 221 GlyGlyLysAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLys 240
Db 891 GGTGGAAAGCCATAGAACATTTTGGCCAAACAGAGAAATAGATTTCAATTTGACATCAA 950
QY 241 ProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260
Db 951 CTCTCCCTGCATCATGCTAAATAATGATATTTCTTTTACGTGACCTTCAACAGCTTACT 1010
QY 261 AspLysIleIleMetLysLysGlyLysGlyGlyIleGlyLysGlyGlnIleLeuSer 280
Db 1011 GATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1070
QY 281 SerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCysHisLeuValLysArg 300
Db 1071 TCAGACACAGACATTTGCTGCCACAGTACAGACACAAATGCGATGCTTGTGAAAAGA 1130
QY 301 ThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuProGlnAsnAsnAlaVal 320
Db 1131 ACAATCGGGCTATTTCTTTTGTGAAGCAGAGAGCTGTGTACCTCAAAATATATCAGTA 1190
QY 321 LeuValAlaSerGlyValAlaSerAsnPhenThrIleArgArgAlaLeuGlnIleLeu 340

Db 1191 CTGTTGCACTTGTGTGTGTGTCGAGTAACTTCTATATCCGACAGCTCTGGAATTTTA 1250
QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProAlaArgLeuCysThrAspAsnGly 360
Db 1251 ACAGAGCAACACAGTGCATTTGTTGTGTCTCTCCAGACATATGACATGATATGCG 1310
QY 361 IleMetIleAlaThrPheAsnGlyIleGlyArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380
Db 1311 ATATGATTCACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1370
QY 381 IleGluGlyIleArgGlyArgGlyProLysCysProLeuGlyValAspIleSerLysGlyVal 400
Db 1371 ATAGAGGCAATCCGCTATGAAACCAAAATGTCTCTTGGAGTACATATGCAAAAGATT 1430
QY 401 GlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 414
Db 1431 GAGAAAGCTTCCATTAAGATCAACATTAATTAATTAATTAATTAATTAATTAATTAAT 1472

RESULT 3
AR541929 1416 bp DNA linear PAT 08-OCT-2004
LOCUS
DEFINITION
ACCESSION AR541929
VERSION AR541929.1 GI:53934009
KEYWORDS
SOURCE
ORGANISM
Unknown.
Unclassified.

REFERENCE
1 (bases 1 to 1416)
Tang, Y.T., Zhou, P., Goodrich, R., Liu, C., Asundi, V., Ren, F.,
Zhang, J., Zhao, Q.A., Yang, Y., Xue, A.J., Wehrman, T., Wang, J.-R.,
Wang, D. and Drmanac, R.T.
Nucleic acids and polypeptides
Patent: US 6743619-A 177 01-JUN-2004;
Location/Qualifiers
1..1416
/organism="unknown"
/mol_type="genomic DNA"

TITLE
JOURNAL
FEATURES
source

ORIGIN
Alignment Scores:
Pred. No.: 0 Length: 1416
Score: 340.00 Matches: 340
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 82.13% Indels: 0
DB: 6 Gaps: 0

US-10-649-273-2 (1-414) x AR541929 (1-1416)
QY 75 ThrGlyGlyIleValAlaProPheAlaAlaGlnGlnLeuHisArgGlyLeuHisLeuArgIle 94
Db 283 ACAGGAGGATGTTCTTCTCAGAGGCTCAACAGCTTCAACAGAAATATTAACAGATTA 342
QY 95 ValGlnGluAlaLeuSerAlaSerLysSerProSerAspSerLeuSerAlaIleAlaThr 114
Db 343 GTACAGAAAGCTTTTCTCAGAGGAGTCTTCAAGTACCTTCAAGAAATGCAACT 402
QY 115 ThrIleLysProGlyLeuAlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeu 134
Db 403 ACCATTAACACAGAGCTTCTTAAACCTGAGAGTGGCTTATCATTTAGCTTACAGCTG 462
QY 135 ValGlyGlnLeuLysLysProPheIleProIleHisHisMetGluAlaHisAlaLeuThr 154
Db 463 GTAGAGCAGATTAAAAAGCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 522
QY 155 IleArgLeuThrAsnLysValGluPheProPheLeuValLeuLeuIleSerGlyIleHis 174
Db 523 ATTAGGTGACCAATTAATTAATTAATTTCTTTTGTGATTTCTTGAAGGTGAC 582
QY 175 CysLeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAsp 194

```
Db      583 TGTCGTGGCATTAGTCAAGAGTTTCAGATTTTCGCTTCTTGAAAGTCTTGGAC 642
Qy      195 ILeAlaProGlyAspMetLeuAspIysValAlaArgArgLeuSerLeuIleIysHisPro 214
Db      643 ATAGCACCAAGTGACATGCTTGACCAAGGTGGCAAGAGAGCTTCTTAATAAACAATCCA 702
Qy      215 GluCysSerThrMetSerGlyGlyIysAlaIleGluHisLeuAlaIysGlnGlyAsnArg 234
Db      703 GAGTGCTCCACCAAGAGTGGTGGGAAAGCCATAGAACATTTGGCCAAACAGAAATAGA 762
Qy      235 PheHisPheAspIleIysProProLeuHisHisAlaIysAsnCysAspPheSerPheThr 254
Db      763 TTTCAATTTTGACATCAAACTCCCTTGCAATGCTAAATAAATTGCAATTTTCTTTACT 822
Qy      255 GlyLeuGlnHisValThrAspIysIleIleMetIysGlyGluGlyGlnGlyIleGlu 274
Db      823 GGACTTCACAACGTTACTGATTAATAATTAATGAAGAAAGAGAGAGATATTGAG 882
Qy      275 LysGlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAla 294
Db      883 AAGGGCAAAATCGTCTTCAGCAGACGATTTGCTGCCACAGTACAGCACACATGGCA 942
Qy      295 CysHisLeuValIysArgThrHisArgAlaIleLeuPheCysIysGlnArgAspLeu 314
Db      943 TGTCACTCTGTGAAGAAGACACATCGGGCTATTCTGTTTGTAGCAGAGAGACTGTTA 1002
Qy      315 ProGlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheTyrIleArg 334
Db      1003 CCTCAAAATTAATGCAGTACTGCTGTCATCTGGTGTGCAAGTAACTTATATATCCGC 1062
Qy      335 ArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArg 354
Db      1063 AGAGCTCTGGAATTTTAACAACCAACGCAACAGTGCATTTGTGTGCTCTCCACAG 1122
Qy      355 LeuCysThrAspAsnGlyIleMetIleAlaTyrPasnGlyIleGluArgLeuArgAlaGly 374
Db      1123 CTATGCACTGATTAAGCATTTGATGCAATGGAATGATTTGAAAGACTACCTGCTGGC 1182
Qy      375 LeuGlyIleLeuHisAspIleGlnGlyIleArgTyrGlnProIysCysProLeuGlyVal 394
Db      1183 TTGGGCATTTTTCATGACATAGACATAGGCAATCCGCTATGAACCAAAATGTCTCTTGAGATA 1242
Qy      395 AspIleSerIysGlnValGlyGlnAlaSerIleIysValProGlnLeuIysMetGluIle 414
Db      1243 GACATATCAAAAGAGTGGAGAAAGCTTCATATAAAGTACCACAATTAATAATGAGATA 1302
```

Search completed: November 11, 2005, 05:46:33
Job time : 7527 secs

CC ameliorating a medical condition, particularly an immune disorder, an
CC aberrant glutamate transport or motor neuron disorder, such as
CC amyotrophic lateral sclerosis (ALS), its juvenile form or an ALS-like
CC condition. The compositions and methods are also useful for diagnosing,
CC prognosticating, treating, ameliorating and/or treating disorders
CC associated with MP-1 activity, e.g. diabetes, cancer, reproductive
CC disorders (e.g. Klinefelter's syndrome, genital warts, or germinal cell
CC aplasia), metabolic disorders (e.g. premature puberty, Kallman syndrome,
CC or Cushing's syndrome), neurodegenerative diseases (Alzheimer's disease,
CC Parkinson's disease, Huntington's disease or Tourette syndrome), liver
CC and renal diseases and immune disorders (e.g. AIDS, rheumatoid arthritis
CC or sepsis), pulmonary diseases (e.g. pneumonia, emphysema or cystic
CC fibrosis) and vascular, inflammatory and neurological disorders (e.g.
CC Alzheimer's disease or Parkinson's disease). This is the amino acid
CC sequence of a metalloprotease MPI protein
XX
SQ Sequence 267 AA:

Query Match 100.0%; Score 1385; DB 5; Length 267;
Best Local Similarity 100.0%; Pred. No. 1.4e-147;
Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEAHALTRLTNKNVEPPLVLLISGGHCLALVQGVSDPFLILGKSLDIPAGMDLKVARR 60
Db 1 MEAHALTRLTNKNVEPPLVLLISGGHCLALVQGVSDPFLILGKSLDIPAGMDLKVARR 60
QY 61 LSLIKPECSMTSGGKAIENHLAKQGNRFPDIKPLHAKNCDPESFTGLQHTVDKIMKK 120
Db 61 LSLIKPECSMTSGGKAIENHLAKQGNRFPDIKPLHAKNCDPESFTGLQHTVDKIMKK 120
QY 121 EKEBIEKGQILSSADIAATVQHTMACHLVKTRTRAILFCQKRDLLPQNNAVLVASGV 180
Db 121 EKEBIEKGQILSSADIAATVQHTMACHLVKTRTRAILFCQKRDLLPQNNAVLVASGV 180
QY 181 ASNFYIRRALIELTNATQCTLLCPPRLCTDNGIMIANNGIERLRAGLGIHIDIGIRYE 240
Db 181 ASNFYIRRALIELTNATQCTLLCPPRLCTDNGIMIANNGIERLRAGLGIHIDIGIRYE 240
QY 241 PKCPLGVDSKEVGEASIKVPOLKMEI 267
Db 241 PKCPLGVDSKEVGEASIKVPOLKMEI 267

RESULT 2
ABG96478
ID ABG96478 standard; protein; 414 AA.
XX
AC ABG96478;
XX
DT 11-DEC-2002 (first entry)
XX
DE Novel human metalloprotease MPI.
XX
XX Metalloprotease; MP-1; immune disorder; glutamate transport; cancer;
XX motor neuron disorder; amyotrophic lateral sclerosis; ALS; diabetes;
XX reproductive disorder; Klinefelter's syndrome; germinal cell aplasia;
XX genital wart; metabolic disorder; premature puberty; Kallman syndrome;
XX Cushing's syndrome; neurodegenerative disease; Alzheimer's disease;
XX Parkinson's disease; Huntington's disease; Tourette syndrome; sepsis;
XX liver disease; renal disease; immune disorder; rheumatoid arthritis;
XX acquired immunodeficiency syndrome; AIDS; pulmonary disease; pneumonia;
XX emphysema; cystic fibrosis; vascular disorder; inflammatory disorder;
XX neurological disorder.
XX
XX Homo sapiens.
XX OS
XX PN MO200272751-A2.
XX
XX PD 19-SEP-2002.
XX
XX PF 05-FEB-2002; 2002WO-US003353.
XX
XX PR 05-FEB-2001; 2001US-026651BP.

PR 10-APR-2001; 2001US-0282814P.
XX
XX (BRIM) BRISTOL-MYERS SQUIBB CO.
XX
XX Chen J, Feder J, Nelson TC, Ducloux F, Krystek S;
PI WPI, 2002-723329/78.
XX
XX N-PSDB; ABS76635.
DR
DR New isolated nucleic acid encoding MP-1 protein, useful for preventing,
PT treating, or ameliorating diseases associated with aberrant
PT metalloproteinase activity, e.g. immune, metabolic, inflammatory and
PT neurological disorders.
XX
XX Claim 5; Fig 1A-C; 473bp; English.
XX
XX The invention describes an isolated nucleic acid molecule (I) encoding a
XX metalloprotease (MP-1). (I) is useful for preventing, treating, or
XX ameliorating a medical condition, particularly an immune disorder, an
XX aberrant glutamate transport or motor neuron disorder, such as
XX amyotrophic lateral sclerosis (ALS), its juvenile form or an ALS-like
XX condition. The compositions and methods are also useful for diagnosing,
XX prognosticating, treating, ameliorating and/or treating disorders
XX associated with MP-1 activity, e.g. diabetes, cancer, reproductive
XX disorders (e.g. Klinefelter's syndrome, genital warts, or germinal cell
XX aplasia), metabolic disorders (e.g. premature puberty, Kallman syndrome,
XX or Cushing's syndrome), neurodegenerative diseases (Alzheimer's disease,
XX Parkinson's disease, Huntington's disease or Tourette syndrome), liver
XX and renal diseases and immune disorders (e.g. AIDS, rheumatoid arthritis
XX or sepsis), pulmonary diseases (e.g. pneumonia, emphysema or cystic
XX fibrosis) and vascular, inflammatory and neurological disorders (e.g.
XX Alzheimer's disease or Parkinson's disease). This is the amino acid
XX sequence of a metalloprotease MPI protein
SQ Sequence 414 AA:

Query Match 100.0%; Score 1385; DB 5; Length 414;
Best Local Similarity 100.0%; Pred. No. 2.8e-147;
Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEAHALTRLTNKNVEPPLVLLISGGHCLALVQGVSDPFLILGKSLDIPAGMDLKVARR 60
Db 148 MEAHALTRLTNKNVEPPLVLLISGGHCLALVQGVSDPFLILGKSLDIPAGMDLKVARR 207
QY 61 LSLIKPECSMTSGGKAIENHLAKQGNRFPDIKPLHAKNCDPESFTGLQHTVDKIMKK 120
Db 208 LSLIKPECSMTSGGKAIENHLAKQGNRFPDIKPLHAKNCDPESFTGLQHTVDKIMKK 267
QY 121 EKEBIEKGQILSSADIAATVQHTMACHLVKTRTRAILFCQKRDLLPQNNAVLVASGV 180
Db 268 EKEBIEKGQILSSADIAATVQHTMACHLVKTRTRAILFCQKRDLLPQNNAVLVASGV 327
QY 181 ASNFYIRRALIELTNATQCTLLCPPRLCTDNGIMIANNGIERLRAGLGIHIDIGIRYE 240
Db 328 ASNFYIRRALIELTNATQCTLLCPPRLCTDNGIMIANNGIERLRAGLGIHIDIGIRYE 387
QY 241 PKCPLGVDSKEVGEASIKVPOLKMEI 267
Db 388 PKCPLGVDSKEVGEASIKVPOLKMEI 414

RESULT 3
ABB05481
ID ABB05481 standard; protein; 414 AA.
XX
XX ABB05481;
XX
XX 19-APR-2002 (first entry)
XX
XX Human O-sialoglycoproteinase-like protein SEQ ID NO.2.
XX
XX Human; O-sialoglycoproteinase-like protein; OSGPUP; enzyme.
XX

OS Homo sapiens.
XX CN1318550-A.
XX
XX 24-OCT-2001.
XX
XX 19-APR-2000; 2000CN-00106834.
XX
XX 19-APR-2000; 2000CN-00106834.
XX
XX (SHAN-) SHANGHAI SHENGYUAN GENE DEV CO LTD.
XX
XX Mao Y, Xie Y;
XX
XX WPI; 2002-115090/16.
XX
XX N-PSDB; ABA93268.
XX
XX O-sialoglycoproteinase-like protein and encoding polynucleotide, useful
XX for diagnosing, preventing and treating related diseases.
XX
XX Claim 1; Page 30-31 (Disclosure); 38pp; Chinese.
XX
XX The present sequence represents human O-sialoglycoproteinase-like protein
XX (OSGPLP). The present invention also describes: (1) the preparation of
XX the OSGPLP protein; (2) applying the OSGPLP protein in diagnosis; (3) the
XX prevention and/or treatment of related diseases; (4) utilizing the OSGPLP
XX protein in screening its agonist, excitomotor and inhibitor and preparing
XX an antibody against the OSGPLP protein; and (5) the use of the OSGPLP
XX polynucleotide sequences, proteins, agonists, excitomotors, inhibitors
XX and antibodies in treating diseases related to the abnormal OSGPLP gene
XX and in preparing the medicine composite for the treatment
XX
XX Sequence 414 AA;

Query Match 100.0%; Score 1385; DB 5; Length 414;
Best Local Similarity 100.0%; Pred. No. 2.8e-147;
Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEAHLTITLTKVBPFLVLLISGHCLLAVGVSPFLIGKSLDIARQMDLVARR 60
DB 148 MEAHLTITLTKVBPFLVLLISGHCLLAVGVSPFLIGKSLDIARQMDLVARR 207
QY 61 LSLIKHECSTMSGKAIIEHLAKOGNRFPHDIKPLHLAKNDFEFTGLOHTDKIMRK 120
DB 208 LSLIKHECSTMSGKAIIEHLAKOGNRFPHDIKPLHLAKNDFEFTGLOHTDKIMRK 267
QY 121 EKEEGIEKGQILSSAADIAATVQHTMACHLVKRTTRAILFCRKORDLLPQNNAVLVASGV 180
DB 268 EKEEGIEKGQILSSAADIAATVQHTMACHLVKRTTRAILFCRKORDLLPQNNAVLVASGV 327
QY 181 ASNVYIRRALEITLNAOTCTLLCPPRLCTNGMIANNGIFRLAAGLIDIGIRYE 240
DB 328 ASNVYIRRALEITLNAOTCTLLCPPRLCTNGMIANNGIFRLAAGLIDIGIRYE 387
QY 241 PKCPGLVDISKVGEVASIKVPOLKMEI 267
DB 388 PKCPGLVDISKVGEVASIKVPOLKMEI 414

RESULT 4
ABJ26654
ID ABJ26654 standard; protein; 414 AA.
XX
XX ABJ26654;
XX
XX 01-MAY-2003 (first entry)
XX
XX Human protein modification + maintenance molecule protein SEQ ID No 8.
XX
XX Cytostatic; antiarteriosclerotic; anti-HIV; antiallergic; anticonvulsant;
XX cerebroprotective; antiparkinsonian; nootropic; antiinflammatory;
XX antitumor; hepatotropic; gynecological; antibacterial; virucide;
XX protozoacide; antiparasitic; cell proliferative disease; PMOD;

KW protein modification and maintenance molecule; immunogenic fragment;
KW cancer; autoimmune; inflammatory disease; neurological disorder;
KW gastrointestinal; developmental; vesicle trafficking disorder; infection;
KW protein-protein interaction; drug-target interaction;
KW gene expression profile; human.

OS Homo sapiens.
XX
XX WO2003000844-A2.
XX
XX 03-JAN-2003.
XX
XX 18-JUN-2002; 2002WO-US019360.
XX
XX 22-JUN-2001; 2001US-0300508P.
XX 06-JUL-2001; 2001US-0303445P.
XX 13-JUL-2001; 2001US-0305405P.
XX 09-AUG-2001; 2001US-0311442P.
XX 24-AUG-2001; 2001US-0314821P.
XX 29-AUG-2001; 2001US-0315992P.
XX 03-MAY-2002; 2002US-0378205P.
XX
XX (INCY-) INCYTE GENOMICS INC.
XX
XX Gandhi AR, Kable AE, Swarnakar A, Hafalia AJA, Tran B, Duggan BM;
XX Warren BA, Ison CH, Honchell CD, Nguyen DB, Lu DAM, Lee BA, Yue H;
XX Forrethe JF, Barroso I, Ramkumar J, Griffin JA, Li JX, Yang J;
XX Thangavelu K, Gietzen KJ, Ding L, Baughn MR, Borowaky ML, Yao MG;
XX Walla NK, Mason PM, Guturjan R, Lee S, Becha SD, Lee ST, Tran UK;
XX Elliott VS, Luo W, Sprague WW, Tang YT, Lu Y, Zebardjian Y;
XX
XX WPI; 2003-184039/18.
XX
XX N-PSDB; ABR23207.
XX
XX New isolated human PMOD polypeptide and polynucleotide, useful for
XX diagnosing, treating and preventing diseases or conditions associated
XX with the aberrant PMOD expression, e.g. cancer, AIDS, atherosclerosis and
XX infections.
XX
XX Claim 63; Page 182-183; 225pp; English.

XX The invention relates to an isolated polypeptide comprising: any of 28
XX sequences of 48-1256 amino acids; a natural amino acid sequence at least
XX 90% identical to the 28 amino acid sequences, 94% identical to a sequence
XX of 703 or 267 amino acids, 96% identical to a sequence of 414 amino
XX acids, or 97% identical to a sequence of 242 amino acids, all given in
XX the specification; or a biologically active or immunogenic fragment of
XX the isolated polypeptide. The polypeptides and polynucleotides are useful
XX in diagnosing, treating and preventing diseases or conditions associated
XX with the decreased expression of protein modification and maintenance
XX molecules (PMOD), such as cell proliferative diseases (e.g. cancer,
XX atherosclerosis), autoimmune/inflammatory diseases (e.g. AIDS,
XX allergies), neurological disorders (e.g. stroke, Parkinson's disease,
XX epilepsy), gastrointestinal (e.g. ulcer, cirrhosis), reproductive (e.g.
XX endometriosis), developmental, vesicle trafficking disorders, and
XX infections (e.g. bacterial, viral, parasitic, protozoal). These are also
XX useful in assessing the effects of exogenous compounds on the expression
XX of nucleic acid and amino acid sequences of PMOD. The PMOD or its
XX fragments are useful in screening compounds for effectiveness as agonist
XX or antagonist of the polypeptides, or in altering the expression of the
XX target polynucleotide and compounds that specifically bind to or modulate
XX the activity of the polypeptide. The microarray is useful in monitoring
XX or measuring protein-protein interactions, drug-target interactions, and
XX gene expression profiles. This sequence represents a human PMOD protein
XX of the invention
XX

SO Sequence 414 AA;

Query Match 100.0%; Score 1385; DB 6; Length 414;
Best Local Similarity 100.0%; Pred. No. 2.8e-147;
Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MEAHLTITLTKVBPFLVLLISGHCLLAVGVSPFLIGKSLDIARQMDLVARR 60

Db 148 MEAHLTRITLNKVEPFLVLLISGHCILALVQGVSDPFLGKSLDIPGMDLKVARR 207
Qy 61 LSLIKHPECSTMSGKAIEHLAKOGRFHPDIKPLHAKNCDPFTGLOHTYDKIMKK 120
Db 208 LSLIKHPECSTMSGKAIEHLAKOGRFHPDIKPLHAKNCDPFTGLOHTYDKIMKK 267
Qy 121 EKEGIEKGQILSSAADIAATVQHTMACHLVKTRTHAILFCQKORDLLPQNNAVLVASGV 180
Db 268 EKEGIEKGQILSSAADIAATVQHTMACHLVKTRTHAILFCQKORDLLPQNNAVLVASGV 327
Qy 181 ASNFYIRALLETITNATQCTLLCPPLRCTDNGIMIANNGIERLRAGLGIHIDIGIRYE 240
Db 328 ASNFYIRALLETITNATQCTLLCPPLRCTDNGIMIANNGIERLRAGLGIHIDIGIRYE 387
Qy 241 PKCPLGVDISKVEGASIKVPOLKMEI 267
Db 388 PKCPLGVDISKVEGASIKVPOLKMEI 414

RESULT 5

AAE29234 standard; proein; 414 AA.

AAE29234;

27-JAN-2003 (first entry)

Human glycoprotease 28472 protein.

Human; adenosine deaminase; seven transmembrane domain receptor; cancer;
7TM; glycoprotease; immune disorder; IGA deficiency; allergy; arhythmia;
rheumatoid arthritis; diabetes; atherosclerosis; cardiovascular disorder;
hyperextension; ischaemic heart disease; obesity; myocardial infarction;
endothelial cell disorder; Grave's disease; psoriasis; brain disorder;
Parkinson's disease; Alzheimer's disease; haemotopoietic disorder;
cerebral oedema; metabolic disorder; liver disorder; platelet disorder;
chromosome mapping; tissue typing; gene therapy; neuroprotective;
cytostatic; anorectic; cardiant; haemostatic.

Homo sapiens.

Location/Qualifiers

1. .108 /note= "Non-transmembrane domain; N-terminal cytoplasmic domain"

38. .369 /note= "Endopeptidase O-sialoglycoprotein hydrolase metalloprotease zinc glycoprotease sialoglycoprotease domain"

109. .132 /note= "Transmembrane domain"

133. .164 /note= "Non-transmembrane domain; non-cytoplasmic loop"

138. .152 /note= "Glycoprotease domain"

165. .189 /note= "Transmembrane domain"

190. .116 /note= "Transmembrane domain"

317. .333 /note= "Non-transmembrane domain; cytoplasmic domain"

334. .414 /note= "Transmembrane domain"

374. .414 /note= "Non-transmembrane domain"

/note= "Sialoglycoprotease type domain"

W0200274960-A2.

26-SEP-2002.

08-NOV-2001; 2001WO-US051427.

XX

PR 08-NOV-2000; 2000US-0246768P.
PR 08-NOV-2000; 2000US-0246772P.
PR 15-NOV-2000; 2000US-0249185P.

(MILL-) MILLENNIUM PHARM INC.

Leiby KR, Kapeller-Libermann R, Glucksmann M;

WPI; 2002-759898/82.

N-PSDB; AAD46856.

New 38650, 28472, 5495, 65507, 81588 or 14354 nucleic acid molecules,

useful for diagnosing and treating cancer, immune, cardiovascular,

hematopoietic, brain, pain, metabolic, liver or platelet disorders, and

in pharmacogenomics.

Claim 1; Fig 8; 178pp; English.

The present invention relates to novel 38650, 28472, 5495, 65507, 81588

or 14354 nucleic acid molecules which encode adenosine deaminase, glyco-

protease or seven transmembrane domain (7TM) receptor family members.

Sequences of the invention are useful in diagnosing and treating cancer

or aberrant cellular proliferation and/or differentiation (e.g. colon or

lung cancer), immune disorders (e.g. selective IGA deficiency, rheumatoid

arthritis, diabetes, allergy), heart or cardiovascular disorders (e.g.,

hypertension, atherosclerosis, arhythmias, ischaemic heart disease,

myocardial infarction, thrombus) including endothelial cell disorders

(e.g. psoriasis, Grave's disease), haematopoietic disorders, brain

disorders (e.g. cerebral oedema, Parkinson's or Alzheimer's disease),

pain and metabolic disorders (e.g. obesity), liver disorders or platelet

disorders. They are also useful in screening assays, predictive medicine

(e.g. diagnostic assays, prognostic assays, monitoring clinical trials

and pharmacogenetics) and prophylactic and therapeutic methods. The

nucleic acids may also be used in chromosome mapping, tissue typing and

forensic biology and as surrogate markers. Sequences of the invention are

also used in gene therapy. The present sequence is human glycoprotease

28472 protein

Sequence 414 AA;

Query Match 97.3%; Score 1348; DB 5; Length 414;

Best Local Similarity 97.4%; Pred. No. 4.3e-143;

Matches 260; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

1 MEAHLTRITLNKVEPFLVLLISGHCILALVQGVSDPFLGKSLDIPGMDLKVARR 60

148 MEAHLTRITLNKVEPFLVLLISGHCILALVQGVSDPFLGKSLDIPGMDLKVARR 207

61 LSLIKHPECSTMSGKAIEHLAKOGRFHPDIKPLHAKNCDPFTGLOHTYDKIMKK 120

208 LSLIKHPECSTMSGKAIEHLAKOGRFHPDIKPLHAKNCDPFTGLOHTYDKIMKK 267

121 EKEGIEKGQILSSAADIAATVQHTMACHLVKTRTHAILFCQKORDLLPQNNAVLVASGV 180

268 EKEGIEKGQILSSAADIAATVQHTMACHLVKTRTHAILFCQKORDLLPQNNAVLVASGV 327

181 ASNFYIRALLETITNATQCTLLCPPLRCTDNGIMIANNGIERLRAGLGIHIDIGIRYE 240

328 ASNFYIRALLETITNATQCTLLCPPLRCTDNGIMIANNGIERLRAGLGIHIDIGIRYE 387

241 PKCPLGVDISKVEGASIKVPOLKMEI 267

388 PKCPLGVDISKVEGASIKVPOLKMEI 414

RESULT 6

ABG71161 standard; proein; 414 AA.

ABG71161;

30-JAN-2003 (first entry)

XX

Accession	Protein Name	Location/Qualifiers
DE	Novel human glycoprotease 28472.	
XX	Cancer; aberrant cell proliferation; aberrant cell differentiation;	
KM	breast cancer; ovarian cancer; prostate cancer; colon cancer;	
KM	lung cancer; immune disorder; heart disorder; cardiovascular disorder;	
KM	endothelial disorder; hematopoietic disorder; blood vessel disorder;	
KM	brain disorder; pain; metabolic disorder; liver disorder; diabetes;	
KM	platelet disorder; carcinoma; sarcoma; leukemia; Hodgkin's disease;	
KM	autoimmune disorder; hypertension; atherosclerosis; heart failure;	
KM	myocardial infarction; ischemic heart disease; Crohn's disease;	
KM	Grave's disease; Kawasaki syndrome; Raynaud's disease; aneurysm;	
KM	cerebral ischemia; peripheral neuropathy; Alzheimer's disease;	
KM	Parkinson's disease; anorexia nervosa; cachexia; glycoprotease; enzyme.	
XX		
OS	Homo sapiens.	
XX		
XX		
FT	Key	Location/Qualifiers
FT	Domain	138..152
FT		/label= Glycoprotease_domain
XX		
XX	W0200277233-A2.	
XX		
PD	03-OCT-2002.	
PF	08-NOV-2001; 2001WO-US046724.	
XX		
XX	08-NOV-2000; 2000US-0246768P.	
PR	08-NOV-2000; 2000US-0246772P.	
PR	15-NOV-2000; 2000US-0249185P.	
XX		
PA	(MILL-) MILLERITUM PHARM INC.	
XX		
PI	Leibyl KR, Kapeller-Libermann R, Gluckmann M;	
XX		
DR	WPI: 2003-023938/02.	
DR	N-PSDB: ABS57020.	
PT	New adenosine deaminase, glycoprotease and seven transmembrane domain	
PT	nucleic acids and polypeptides, designated 38650, 28472, 5495, 65507,	
PT	81588 and 14354, useful for treating e.g. leukemias, Hodgkin's disease or	
PS	hypertension.	
XX		
PS	Claim 4; Fig 8A-B; 178bp; English.	
XX		
CC	The invention describes isolated 38650, 28472, 5495, 65507, 81588 and	
CC	14354 nucleic acid molecules (I) and their encoded polypeptides (II). The	
CC	38650 nucleic acid molecule comprises a sequence encoding adenosine	
CC	deaminase. The 28472 nucleic acid molecule comprises a sequence encoding	
CC	a human glycoprotease. The 5495, 65507, 81588 and 14354 comprise	
CC	sequences that encode a human seven transmembrane domain (7TM). The	
CC	38650, 28472, 5495, 65507, 81588 or 14354 nucleic acid and polypeptide	
CC	sequences are useful for diagnosing, preventing or treating a subject	
CC	with or at risk of developing a disorder, e.g. cancer or aberrant	
CC	cellular proliferation and/or differentiation (e.g. breast, ovarian,	
CC	prostate, colon or lung cancer), immune disorders, heart disorders,	
CC	cardiovascular disorders, endothelial disorders, hematopoietic disorders,	
CC	blood vessel disorders, brain disorders, pain and metabolic disorders,	
CC	liver disorders or platelet disorders. These disorders include carcinoma,	
CC	sarcoma, leukaemia, Hodgkin's disease, autoimmune disorders,	
CC	hypertension, atherosclerosis, heart failure, myocardial infarction,	
CC	ischemic heart disease, Crohn's disease, Grave's disease, Kawasaki	
CC	syndrome, Raynaud's disease, aneurysm, cerebral ischaemia, peripheral	
CC	neuropathy, Alzheimer's disease, Parkinson's disease, anorexia nervosa,	
CC	cachexia or diabetes. This is the amino acid sequence of the novel human	
CC	glycoprotease 28472	
XX		
SO	Sequence 414 AA;	
XX		
XX	Query Match 97.3%; Score 1348; DB 6; Length 414;	
XX	Best local similarity 97.4%; Pred. No. 4.3e-143;	
XX	Matches 260; Conservative 3; Mismatches 4; Indels 0; Gaps 0	

Db	148	MEAAALITRLTNKPEPFVLLISGGICLALVGVSDPFLLGSLDIAQDMIDKARR	207
Qy	61	LSLIKHECSTMSGGKALIEHLAKQGNRFHEDIKPLHLHAKNCDFSPTGLQHTDKITMKK	120
Db	208	LSLIKHECSTMSGGKALIEHLAKQGNRFHEDIKPLHLHAKNCDFSPTGLQHTDKKNENR	267
Qy	121	EKEGIEKGQLSSAADIATVQHTMACHLVKRTHRALILPKCKNDLLPNNNAVIVASGV	180
Db	268	KQEGIEKGQLSSAADIATVQHTMACHLVKRTHRALILPKCKNDLLPNNNAVIVASGV	327
Qy	181	ASNFYIRALAEILTNATQCTLLCPPLCTDNGIMIMANGIERLRAGILHDIEGIRYE	240
Db	328	ASNFYIRALAEILTNATQCTLLCPPLCTDNGIMIMANGIERLRAGILHDIEGIRYE	387
Qy	241	PKCPLGVDISKVGEASIKVPOLMEI	267
Db	388	PKCPLGVDISKVGEASIKVPOLMEI	414
RESULT 7			
ID	ABU09569	standard; protein; 414 AA.	
AC	ABU09569;		
DT	08-JUL-2003	(first entry)	
DE	Human glycoprotease encoded by cDNA 28472.		
KM	Human; enzyme; cancer; aberrant cellular proliferation; differentiation;		
KM	immune disorders; heart disorder; brain disorder;		
KM	cardiovascular disorder; endothelial cell disorder; pain disorder;		
KM	haematopoietic disorder; blood vessel disorder; metabolic disorder;		
KM	liver disorder; platelet disorder; glycoprotease.		
OS	Homo sapiens.		
PN	US2003009017-A1.		
PD	09-JAN-2003.		
XX	08-NOV-2001;	2001US-00012140.	
PF	08-NOV-2000;	2000US-0246768P.	
PR	08-NOV-2000;	2000US-0246772P.	
PR	15-NOV-2000;	2000US-0249185P.	
XX	(LEIB/) LEIBY K R.		
PA	(KAPE/) KAPPELLER-LIBERMANN R.		
PA	(GLUC/) GLUCKSMANN M A.		
PI	Leiby KR, Kapeller-Libermann R, Glucksmann MA;		
XX	WPI: 2003-428888/40.		
DR	N-PSDB; ACA60887.		
PT	New isolated 38650, 28472, 5495, 65507, 81588 and 14354 nucleic acid		
PT	molecules, useful for diagnosing, treating cancer, pain, or immune,		
PT	heart, endothelial cell, hematopoietic, blood vessel, brain, metabolic		
PT	and liver disorders.		
XX	Claim 4; Fig 8; 90pp; English.		
XX	The invention relates to an isolated 38650 (encoding adenosine		
CC	deaminase), 28472 (encoding a glycoprotease), 5495 (encoding a 7		
CC	transmembrane receptor), 65507, 81588 or 14354 nucleic acid molecule (or		
CC	a sequence which is at least 60% identical to the six nucleic acids or		
CC	their open reading frames, fragments of at least 15 nucleotides,		
CC	naturally occurring variants, or a DNA insert of the plasmid deposited		
CC	with the American Type Culture Collection as Accession No. not defined in		
CC	the specification, which encodes the amino acid sequence). Also included		
CC	are a host cell containing the nucleic acids (used to produce the		

CC proteins), the encoded proteins, an antibody that selectively binds to
CC the polypeptide, and identifying a compound that binds to/modulates the
CC activity of the polypeptide. The nucleic acid molecules, polypeptides and
CC methods are useful for diagnosing, treating cancer, aberrant cellular
CC proliferation and/or differentiation, immune disorders, heart disorders,
CC cardiovascular disorders including endothelial cell disorders,
CC hematopoietic disorders, blood vessel disorders, brain disorders, pain
CC and metabolic disorders, liver disorders and platelet disorders (many
CC examples of these disorders are given in the specification). The present
CC sequence is the Human glycoproteinase encoded by cDNA 28472

XX Sequence 414 AA;

Query Match 97.3%; Score 1348; DB 6; Length 414;

Best Local Similarity 97.4%; Pred. No. 4.3e-143;

Matches 260; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 MEAHLTIRLTNKKVEPPLVLLISGHCCLALVQGVSDPFLLGKSLDIAPGMDLDRVAR 60
DB 148 MEAHLTIRLTNKKVEPPLVLLISGHCCLALVQGVSDPFLLGKSLDIAPGMDLDRVAR 207
QY 61 LSLIKHECSTMSGKAI EHLAKGNRPFDIKPLHAKNCDPFTGLQHTVDKIMKK 120
DB 208 LSLIKHECSTMSGKAI EHLAKGNRPFDIKPLHAKNCDPFTGLQHTVDKIMKK 267
QY 121 EKEEGIEKGQILSSAADIATVQHTMACLVKTRTRAILFCRKORDLPONNAVIVASGV 180
DB 268 EKEEGIEKGQILSSAADIATVQHTMACLVKTRTRAILFCRKORDLPONNAVIVASGV 327
QY 181 ASNFYIRRALIELTNATQCTLLCPPRLCTDNGIMIANNGIERLRAGLIIHDSIGRIYE 240
DB 328 ASNFYIRRALIELTNATQCTLLCPPRLCTDNGIMIANNGIERLRAGLIIHDSIGRIYE 387
QY 241 PKCPLGVDISKVEGASIKVPOLKMEI 267
DB 388 PKCPLGVDISKVEGASIKVPOLKMEI 414

RESULT 8

ADA54471

ID ADA54471 standard; protein, 364 AA.

AC ADA54471;

DT 20-NOV-2003 (first entry)

DE Human protein, SEQ ID 2039.

XX Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nocotropic;

KM Gene Therapy; human; secretory protein; membrane proteins; cancer;

KW inflammatory disease; osteoporosis; neurological disease.

OS Homo sapiens.

PN EP1293569-A2.

PD 19-MAR-2003.

PF 21-MAR-2002; 2002EP-00006586.

PR 14-SEP-2001; 2001JP-00328381.

PR 24-JAN-2002; 2002US-0350435P.

PA (HELI-) HELIX RES INST.

PA (REAS-) RES ASSOC BIOTECHNOLOGY.

PI Iisogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;

PI Yamamoto U, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;

PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Maehno Y;

XX WPI; 2003-395539/38.

DR N-PSDB; ADA52832.

XX

PT New polynucleotides encoding full-length polypeptides, e.g. secretory
PT and/or membrane proteins, useful for developing medicines for diseases in
PT which the gene is involved, or as target molecules for gene therapy.

XX Claim 14; SEQ ID NO 2039; 205pp; English.

XX The present invention relates to novel human secretory or membrane
CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
CC ADA54071). The coding sequences are useful in the gene therapy of
CC diseases caused by abnormalities of the proteins, e.g. cancer,
CC inflammatory diseases, osteoporosis or neurological disease.

XX Sequence 364 AA;

Query Match 79.8%; Score 1105; DB 6; Length 364;

Best Local Similarity 98.6%; Pred. No. 1.1e-115;

Matches 214; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MEAHLTIRLTNKKVEPPLVLLISGHCCLALVQGVSDPFLLGKSLDIAPGMDLDRVAR 60
DB 148 MEAHLTIRLTNKKVEPPLVLLISGHCCLALVQGVSDPFLLGKSLDIAPGMDLDRVAR 207
QY 61 LSLIKHECSTMSGKAI EHLAKGNRPFDIKPLHAKNCDPFTGLQHTVDKIMKK 120
DB 208 LPLIKHECSTMSGKAI EHLAKGNRPFDIKPLHAKNCDPFTGLQHTVDKIMKK 267
QY 121 EKEEGIEKGQILSSAADIATVQHTMACLVKTRTRAILFCRKORDLPONNAVIVASGV 180
DB 268 EKEEGIEKGQILSSAADIATVQHTMACLVKTRTRAILFCRKORDLPONNAVIVASGV 327
QY 181 ASNFYIRRALIELTNATQCTLLCPPRLCTDNGIMIA 217
DB 328 ASNFYIRRALIELTNATQCTLLCPPRLCTDNGIMIA 364

RESULT 9

ABB69133

ID ABB69133 standard; protein, 409 AA.

AC ABB69133;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 34191.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US009231.

PR 23-MAR-2000; 2000US-0191637P.

PR 11-JUL-2000; 2000US-00614150.

PA (PEKE) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

DR WPI; 2001-656860/75.

DR N-PSDB; ABL13236.

PT New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions.

XX Disclosure; SEQ ID NO 34191, 21pp + Sequence Listing; English.

PS The invention relates to an isolated nucleic acid detection reagent

XX

PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137724P.
PR 07-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138549P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 23-JUN-1999; 99US-0140595P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140822P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148655P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149930P.
PR 23-AUG-1999; 99US-0150566P.
PR 25-AUG-1999; 99US-0150884P.
PR 26-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0156559P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.

PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 16-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161932P.
PR 28-OCT-1999; 99US-0161933P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 28.8%; Score 398.5; DB 3; Length 245;
Best Local Similarity 38.7%; Pred. No. 5,6e-36;
Matches 98; Conservative 34; Mismatches 90; Indels 31; Gaps 6;

QY 1 MEAHATLRLT-NKVEPFVLVLLISGHCCLALVGVSDPFLILGKSLDIAPGMDLKVAK 59
1 MEATIVARLVEQSLFFPMALLISGCHNLVLAKHQLQYQLGTTVDALGEADKTKAK 60
Db 1
QY 60 RLSIIRKPECSMGSGKAIIEHLAKQGNRFHDIKRLPHAKKCNDSFPGLOHVTDKIIMK 119
61 WIGIDMH-----RSGGPAVEELALEGDAKSVKFNVMKYNKHCNFSYAGLKTQVLAIEA 115
Db 1
QY 120 KEKEGIEKQIILSA-----ADIAATQHTMAGHLYKRTTHAILFCRQRDLIPQNN 171
116 KE-----IDAKCPVSSATNEDRRNRADIASFORVAVLHLEKCEKRAIDMALE---LEPSI 168
Db 1
QY 172 AVLVASGVASNFYIRRLAILITVNAOTCTLLCPRLCTDNGIMIAMNGIERLRAGLIL 231
169 KHWIISGVASNKYRLRLINNIIVENKNIKLVCPPLSLCTDNGVAVMAGLEHFRVG---- 224
Db 1
QY 232 HDIEGIRYEPKCP 244
225 -----RYDPEPP 231
Db 1

RESULT 12
AAG19286
ID AAG19286 standard; protein; 439 AA.

XX AAG19286;

XX 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 21030.

XX Protein identification; signal transduction pathway; metabolic pathway;
XX Hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.

OS Arabidopsis thaliana.

XX EPI033405-A2.
XX 06-SEP-2000.
XX PD
XX 25-FEB-2000; 2000EP-00301439.
XX 25-FEB-1999; 99US-0121825P.
XX 05-MAR-1999; 99US-0123180P.
XX 09-MAR-1999; 99US-0123548P.
XX 23-MAR-1999; 99US-0125788P.
XX 25-MAR-1999; 99US-0126264P.
XX 29-MAR-1999; 99US-0126785P.
XX 01-APR-1999; 99US-0127462P.
XX 06-APR-1999; 99US-0128234P.
XX 08-APR-1999; 99US-0128714P.
XX 16-APR-1999; 99US-0129845P.
XX 19-APR-1999; 99US-0130077P.
XX 21-APR-1999; 99US-0130449P.
XX 23-APR-1999; 99US-0130510P.
XX 23-APR-1999; 99US-0130891P.
XX 28-APR-1999; 99US-0131449P.
XX 30-APR-1999; 99US-0132048P.
XX 30-APR-1999; 99US-0132407P.
XX 04-MAY-1999; 99US-0132484P.
XX 05-MAY-1999; 99US-0132485P.
XX 06-MAY-1999; 99US-0132486P.
XX 06-MAY-1999; 99US-0132487P.
XX 07-MAY-1999; 99US-0132863P.
XX 11-MAY-1999; 99US-0134256P.
XX 14-MAY-1999; 99US-0134218P.
XX 14-MAY-1999; 99US-0134219P.
XX 14-MAY-1999; 99US-0134370P.
XX 14-MAY-1999; 99US-0134371P.
XX 18-MAY-1999; 99US-0134768P.
XX 19-MAY-1999; 99US-0134941P.
XX 20-MAY-1999; 99US-0135124P.
XX 24-MAY-1999; 99US-0135313P.
XX 25-MAY-1999; 99US-0135629P.
XX 27-MAY-1999; 99US-0136021P.
XX 27-MAY-1999; 99US-0136392P.
XX 28-MAY-1999; 99US-0136782P.
XX 01-JUN-1999; 99US-0137222P.
XX 03-JUN-1999; 99US-0137528P.
XX 04-JUN-1999; 99US-0137502P.
XX 07-JUN-1999; 99US-0137724P.
XX 08-JUN-1999; 99US-0138094P.
XX 10-JUN-1999; 99US-0138540P.
XX 10-JUN-1999; 99US-0138847P.
XX 14-JUN-1999; 99US-0139119P.
XX 16-JUN-1999; 99US-0139452P.
XX 16-JUN-1999; 99US-0139453P.
XX 17-JUN-1999; 99US-0139492P.
XX 18-JUN-1999; 99US-0139454P.
XX 18-JUN-1999; 99US-0139455P.
XX 18-JUN-1999; 99US-0139456P.
XX 18-JUN-1999; 99US-0139457P.
XX 18-JUN-1999; 99US-0139458P.
XX 18-JUN-1999; 99US-0139459P.
XX 18-JUN-1999; 99US-0139460P.
XX 18-JUN-1999; 99US-0139461P.
XX 18-JUN-1999; 99US-0139462P.
XX 18-JUN-1999; 99US-0139463P.
XX 18-JUN-1999; 99US-0139750P.
XX 18-JUN-1999; 99US-0139750P.
XX 21-JUN-1999; 99US-0139817P.
XX 22-JUN-1999; 99US-0139899P.
XX 23-JUN-1999; 99US-0140035P.
XX 23-JUN-1999; 99US-0140035P.
XX 24-JUN-1999; 99US-0140035P.
XX 28-JUN-1999; 99US-0140023P.
XX 29-JUN-1999; 99US-0140091P.
XX 30-JUN-1999; 99US-0141287P.

PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 20-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 03-AUG-1999; 99US-0146389P.
PR 04-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 05-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 06-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149425P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.

PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155658P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 28.8%; Score 398.5; DB 3; length 439;
Best Local Similarity 38.7%; Pred. No. 1.3e-35;
Matches 98; Conservative 34; Mismatches 90; Indels 31; Gaps 6;

QY 1 MEAHALTRLT-NKVEPPLVLIIISGRCILALVQGVSPFLIGKSLDIAPEMDIKVAR 59
DB 195 MEAHTLVARLVQGLSPFPFALLISGHNILVLIAHKLCQYTOIGTVDDAIGRAPDKTAK 254
QY 60 RLSLIKHPKPCSTMSGKALIEHLAKQGNRFHFDIKRPLHAKNCDPFTGLQHTDKTIIMK 119
DB 255 WGLDMH-----RSGPVAVEELALEGDASKVKNVPMKHKDCNFSYAGLKTQVRLAIEA 309
QY 120 KEKEGIEKQILISSA-----ADIAATVOHTMACHVKTHTRAILPFCKORLLPORN 171
DB 310 KE-----IDAKCPVSSATINEDRRRADIIAASFQVAVLAHLEBKERRAIDALE---LEPSI 362
QY 172 AVLVASGVASNFYIRALIEILTNATQCTLLCPPLCTDNGIMIAMNGIERLPAGCIL 231
DB 363 KHVVISGVASNKYVALRLNNIIVENKULKIVCPPLCTDNGVMVAMWGIEHFRVG---- 418
QY 232 HDIEGIRYRPKCP 244
DB 419 -----RYDPPPP 425

RESULT 13
AAG19285
ID AAG19285 standard; protein; 444 AA.
XX
AC AAG19285;
XX

PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 28.8%; Score 398.5; DB 3; Length 444;
Best Local Similarity 38.7%; Pred. No. 1.4e-35;
Matches 98; Conservative 34; Mismatches 90; Indels 31; Gaps 6;

QY 1 MEAAHTIRLT-NKVEPFLVLLISGGHCLALVGVSDFLLGKSLDIAPDMDKVAR 59
DB 200 MEAHTLVARLVEQELISPEFMALLISGHNLLVLAHKQCYQLGTTVDALIGBAFDKTA 259
QY 60 RLSTIKPEPSTMSGKAIIEHLAKQGNRFHDIKPELHAKNCDFSPFGLOHVTDKIIMK 119
DB 260 WLGLDMH-----RSGGPAVEELALEGDAKSVKFNPMKTHKDCNFSYAGLKTQVRLAIFA 314
QY 120 KEKEGIEKQILSSA-----ADIATVQHTMACHLVKRTTHAIFLCKQRDLPPONN 171
DB 315 KE-----IDAKCPVSAATNEDRRNRADIASFORVAVLHLEKCEBALDWALE---LEPSI 367
QY 172 AVLVASGCVASNFYIRALIELTNATQCTLLCPPRLCTDNGIMIAWNGIERLRAGLIGL 231
DB 368 KHWYISGCVASNKYVRLRLANNI VENKONKLVCPPSLCTDNGVMAWMTGLEHFRVG---- 423
QY 232 HDIEGIRYEPKPC 244
|||:|:|

DB 424 -----RYDPPP 430
RESUT 14
AAE31054
ID AAE31054 standard; protein; 251 AA.
XX
AC AAE31054;
XX
DT 24-FEB-2003 (first entry)
XX
DE Ehrlichia ruminantium hypothetical o-sialoglycoprotein endopeptidase.
XX
KW Vaccine; immunity; rickettsial infection; spotted fever; heart water;
KW typhus; pathogen; immunostimulant; antibacterial; lhworf4; enzyme;
KW o-sialoglycoprotein endopeptidase.
XX
OS Ehrlichia ruminantium.
XX
PN WO20026652-A2.
XX
PD 29-AUG-2002.
XX
PF 20-FEB-2002; 2002WO-US005772.
XX
PR 20-FEB-2001; 2001US-0269944P.
XX
PA (UFL) UNIV FLORIDA.
XX
PI Barbet AF, Whitmire WW, Kamper SM, Simbi BH, Ganta RR;
PI Moreland AL, Mwangi DM, McGuire TC, Mahan SM;
XX
DR MPI: 2002-723186/78.
XX
DR N-PSDB; AAD48239.
XX
PT New Ehrlichia ruminantium polynucleotides, useful as vaccines for
PT inducing protective immunity, and protecting animals or humans against
PT rickettsial diseases, e.g. typhus, spotted fever or heart water.
XX
PS Claim 8; Page 67-68; 206pp; English.
XX
XX The present invention relates to nucleic acid vaccines for conferring
CC immunity to rickettsial infection, including Ehrlichia ruminantium
CC (formerly Cowdria ruminantium). The invention also relates to novel E.
CC ruminantium polynucleotides and their corresponding proteins. Sequences
CC of the invention are useful for inducing immunity, particularly
CC protective immunity. They are also useful for detecting the presence of
CC E. ruminantium in a biological sample. They are useful in vaccines for
CC protecting animals or humans against rickettsial diseases, e.g. typhus,
CC spotted fever or heart water. Sequences of the invention are useful for
CC detecting antibodies to pathogens. The present sequence is E. ruminantium
CC hypothetical o-sialoglycoprotein endopeptidase (lhworf4) protein
CC
SQ Sequence 251 AA;

QY 1 MEAAHTIRLT-NKVEPFLVLLISGGHCLALVGVSDFLLGKSLDIAPDMDKVAR 60
DB 20 LEAHTLLIMFHDIDPFLVLLISGGHCLALVGVSDFLLGKSLDIAPDMDKVAR 79
QY 61 LSTIKPEPSTMSGKAIIEHLAKQGNRFHDIKPELHAKNCDFSPFGLOHVTDKIIMK 120
DB 80 LNL-GYP-----GCPILIEKKSIMGSKSFILPRALINRLGCDPFGITAVANIIVNQ 132
QY 121 KEKEGIEKQILSSADIATVQHTMACHLVKRTTHAIFLCKQRDLPPONN 175
DB 133 KTIIDN-----DFICNISAFQDCIGDILVNIITNAIHMSQAINCKINK-----LV 177
QY 176 ASGCVASNFYIRALIELTNATQCTLLCPPRLCTDNGIMIAWNGIERLRAGLIGL 235
|||:|:|

Dd	178	VTGGVAAAHLLFNRRISICVKDNNFEVLVPETELCTDNGIMVGAGIENLSKGI-----VS	232
Qy	236	GIRYEPK 242 : 	
Db	233	NLDFVPK 239	
RESULT 15			
ID	ABU22934	standard; protein; 382 AA.	
AC	ABU22934;		
XX			
DT	19-JUN-2003	(first entry)	
DE	Protein encoded by Prokaryotic essential gene #8461.		
XX			
KW	Antisense; prokaryotic essential gene; cell proliferation; drug design.		
OS	Bordetella pertussis.		
XX			
FN	WO200277183-A2.		
PD			
XX	03-OCT-2002.		
PF			
XX	21-MAR-2002; 2002WO-US009107.		
PR			
XX	21-MAR-2001; 2001US-00815242.		
PR	06-SEP-2001; 2001US-00948993.		
PR	25-OCT-2001; 2001US-0342923P.		
PR	08-FEB-2002; 2002US-00072851.		
PR	06-MAR-2002; 2002US-0362699P.		
PA	(ELIT-) ELITRA PHARM INC.		
XX			
F1	Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KU, Zyskind JW;		
F1	Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;		
XX			
DR	WPI; 2003-029926/02.		
XX	N-PSDB; ACA26804.		
PT	New antisense nucleic acids, useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs.		
PS	Claim 25; SEQ ID NO 50858; 1766gp; English.		

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2005, 16:43:58 ; Search time 70.5228 Seconds

(without alignments)
1584.102 Million cell updates/sec

Title: US-10-649-273-2_COPY_148_414

Perfect score: 1385

Sequence: 1 MEAHALTRLTNKVFPPLV.....DISKEVGASIKVQLKMEI 267

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1867879 seqs, 418409474 residues

Total number of hits satisfying chosen parameters: 1867879

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/1/pubppaa/US07_PUBCOMB.pep.*

2: /cgn2_6/ptodata/1/pubppaa/PCT_NEW_PUB.pep.*

3: /cgn2_6/ptodata/1/pubppaa/US06_NEW_PUB.pep.*

4: /cgn2_6/ptodata/1/pubppaa/US06_PUBCOMB.pep.*

5: /cgn2_6/ptodata/1/pubppaa/US07_NEW_PUB.pep.*

6: /cgn2_6/ptodata/1/pubppaa/PCTUS_PUBCOMB.pep.*

7: /cgn2_6/ptodata/1/pubppaa/US08_NEW_PUB.pep.*

8: /cgn2_6/ptodata/1/pubppaa/US08_PUBCOMB.pep.*

9: /cgn2_6/ptodata/1/pubppaa/US09_PUBCOMB.pep.*

10: /cgn2_6/ptodata/1/pubppaa/US09_PUBCOMB.pep.*

11: /cgn2_6/ptodata/1/pubppaa/US09C_PUBCOMB.pep.*

12: /cgn2_6/ptodata/1/pubppaa/US10_NEW_PUB.pep.*

13: /cgn2_6/ptodata/1/pubppaa/US10_PUBCOMB.pep.*

14: /cgn2_6/ptodata/1/pubppaa/US10C_PUBCOMB.pep.*

15: /cgn2_6/ptodata/1/pubppaa/US10C_PUBCOMB.pep.*

16: /cgn2_6/ptodata/1/pubppaa/US10C_PUBCOMB.pep.*

17: /cgn2_6/ptodata/1/pubppaa/US10E_PUBCOMB.pep.*

18: /cgn2_6/ptodata/1/pubppaa/US10_NEW_PUB.pep.*

19: /cgn2_6/ptodata/1/pubppaa/US11_PUBCOMB.pep.*

20: /cgn2_6/ptodata/1/pubppaa/US11_NEW_PUB.pep.*

21: /cgn2_6/ptodata/1/pubppaa/US60_NEW_PUB.pep.*

22: /cgn2_6/ptodata/1/pubppaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1385	100.0	267	US-10-067-443-22	Sequence 22, Appl
2	1385	100.0	267	US-10-649-273-22	Sequence 22, Appl
3	1385	100.0	267	US-10-651-722-22	Sequence 22, Appl
4	1385	100.0	414	US-10-067-443-2	Sequence 2, Appl
5	1385	100.0	414	US-10-649-273-2	Sequence 2, Appl
6	1385	100.0	414	US-10-651-722-2	Sequence 2, Appl
7	1385	100.0	414	US-10-480-988-8	Sequence 2, Appl
8	1358.5	98.1	439	US-10-067-443-19	Sequence 8, Appl
9	1358.5	98.1	439	US-10-649-273-19	Sequence 19, Appl
10	1358.5	98.1	439	US-10-651-722-19	Sequence 19, Appl
11	1348	97.3	414	US-10-012-140-5	Sequence 5, Appl

12	1105	79.8	364	15	US-10-094-749-2039	Sequence 2039, Ap
13	422.5	30.5	445	15	US-10-424-599-209259	Sequence 209259,
14	415.5	30.0	409	20	US-11-097-143-34191	Sequence 34191, A
15	409.5	29.6	463	14	US-10-067-443-3	Sequence 3, Appl
16	409.5	29.6	463	15	US-10-649-273-3	Sequence 3, Appl
17	409.5	29.6	463	15	US-10-651-722-3	Sequence 3, Appl
18	359	25.9	444	16	US-10-437-965-113732	Sequence 113732,
19	342	24.7	251	14	US-10-081-051-9	Sequence 9, Appl
20	321.5	23.2	382	15	US-10-282-122A-50858	Sequence 50858, A
21	317	22.9	333	14	US-10-012-140-25	Sequence 25, Appl
22	314.5	22.7	312	15	US-10-282-122A-44499	Sequence 44499, A
23	308.5	22.3	348	15	US-10-282-122A-63156	Sequence 63156, A
24	304.5	22.0	343	15	US-10-282-122A-67227	Sequence 67227, A
25	294	21.2	347	14	US-10-012-140-24	Sequence 24, Appl
26	291.5	21.0	341	9	US-09-815-242-11798	Sequence 11798, A
27	291.5	21.0	341	15	US-10-282-122A-66200	Sequence 66200, A
28	291.5	21.0	341	18	US-10-958-216-50	Sequence 50, Appl
29	289.5	20.9	341	18	US-10-958-216-52	Sequence 52, Appl
30	289.5	20.9	342	9	US-09-815-242-11043	Sequence 11043, A
31	289.5	20.9	342	15	US-10-282-122A-58204	Sequence 58204, A
32	280.5	20.3	340	15	US-10-282-122A-68438	Sequence 68438, A
33	279.5	20.2	337	9	US-09-815-242-13780	Sequence 13780, A
34	279.5	20.2	337	15	US-10-282-122A-75465	Sequence 75465, A
35	279.5	20.2	337	15	US-10-282-122A-78161	Sequence 78161, A
36	277.5	20.0	335	15	US-10-282-122A-55404	Sequence 55404, A
37	277	20.0	309	15	US-10-282-122A-61538	Sequence 61538, A
38	277	20.0	341	15	US-10-282-122A-67993	Sequence 67993, A
39	276	19.9	421	14	US-10-067-443-4	Sequence 4, Appl
40	276	19.9	421	14	US-10-067-443-28	Sequence 28, Appl
41	276	19.9	421	15	US-10-649-273-4	Sequence 4, Appl
42	276	19.9	421	15	US-10-649-273-28	Sequence 28, Appl
43	276	19.9	421	15	US-10-651-722-4	Sequence 4, Appl
44	276	19.9	421	15	US-10-651-722-28	Sequence 28, Appl
45	275.5	19.9	337	9	US-09-815-242-10304	Sequence 10304, A

ALIGNMENTS

RESULT 1

US-10-067-443-22

Sequence 22, Application US/10067443

Publication No. US20030082782A1

GENERAL INFORMATION:

APPLICANT: Bristol-Myers Squibb Company

TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN

FILE REFERENCE: D0073 NP

CURRENT APPLICATION NUMBER: US/10/067,443

CURRENT FILING DATE: 2002-02-05

PRIOR APPLICATION NUMBER: US 60/266,518

PRIOR FILING DATE: 2001-02-05

PRIOR APPLICATION NUMBER: US 60/282,814

PRIOR FILING DATE: 2001-04-10

NUMBER OF SEQ ID NOS: 71

SOFTWARE: PatentIn version 3.0

SEQ ID NO 22

LENGTH: 267

TYPE: PRT

ORGANISM: homo sapiens

US-10-067-443-22

Query Match 100.0%; Score 1385; DB 14; Length 267;

Best Local Similarity 100.0%; Pred. No. 2.2e-139; Indels 0; Gaps 0;

Matches 267; Conservative 0; Mismatches 0;

QY 1 MEAHALTRLTNKVFPPLVLTISGHCILAVOGSDPLILGKSLDIARGMIDKVAR 60

DB 1 MEAHALTRLTNKVFPPLVLTISGHCILAVOGSDPLILGKSLDIARGMIDKVAR 60

QY 61 LSLIKHPECSITSGKATIEHLAKQGNRPFDIKPILHAQNDPFSFTGQHYTDKIMK 120

DB 61 LSLIKHPECSITSGKATIEHLAKQGNRPFDIKPILHAQNDPFSFTGQHYTDKIMK 120

Qy	121	EKEBEIEKGOLISSADIAATVQHTMACHTVYRTHRAILLFCRCORDILLPONNAVALVMSGGV	180
Db	121	EKEBEIEKGOLISSAADIAATVQHTMACHTVYRTHRAILLFCRCORDILLPONNAVALVMSGGV	180
Qy	181	ASNFIYRRALELLTNAOTCTLLCPPERLCTDNGIMIANNGIERLRAGLGIILHIDIEGIRYE	240
Db	181	ASNFIYRRALBELLTNAOTCTLLCPPERLCTDNGIMIANNGIERLRAGLGIILHIDIEGIRYE	240
Qy	241	PKCPPLGVDISKEVGEASIKVPOLAKMEI	267
Db	241	PKCPPLGVDISKEVGEASIKVPOLAKMEI	267

```

RESULT 2
US-10-649-273-22
: Sequence 22, Application US/10649273
: Publication No. US20040043407A1
GENERAL INFORMATION:
APPLICANT: Bristol-Myers Squibb Company
: TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1-1
: FILE REFERENCE: D0073 CNT
: CURRENT APPLICATION NUMBER: US/10/649,273
: PRIOR FILING DATE: 2003-08-27
: PRIOR APPLICATION NUMBER: US 60/266,518
: PRIOR FILING DATE: 2001-02-05
: PRIOR APPLICATION NUMBER: US 10/067,443
: PRIOR FILING DATE: 2002-02-05
: PRIOR APPLICATION NUMBER: US 60/282,814
: PRIOR FILING DATE: 2001-04-10
: NUMBER OF SEQ ID NOS: 71
: SOFTWARE: PatentIn version 3.2
: SEQ ID NO 22
: LENGTH: 267
: TYPE: PRT
: ORGANISM: homo sapiens
: US-10-649-273-22

```

Query Match	100.0%	Score 1385	DB 15	Length 267
Local Similarity	100.0%	Pred. No. 2.2e-139		
Matches	267	Conservative	0	Mismatches 0; Indels 0; Gaps 0
Qy	1	MEAAHLLTRLTKEVFPFVLVLLISGGHCHLALVQGVSDLLGKSLDIAPGMLDKVARR	60	
Db	1	MEAAHLLTRLTKEVFPFVLVLLISGGHCHLALVQGVSDLLGKSLDIAPGMLDKVARR	60	
Qy	61	LSLKGPECSITMSGKAEIHLAKQGNRFPHDIKPLAHAKNCDFSFTGLQHTYDKIIMKK	120	
Db	61	LSLKGPECSITMSGKAEIHLAKQGNRFPHDIKPLAHAKNCDFSFTGLQHTYDKIIMKK	120	
Qy	121	EKEBGEIEKGQILSSAADIAATVQHTMACHLVETRTIRAILFCQORDILLPNNNAVLVASGV	180	
Db	121	EKEBGEIEKGQILSSAADIAATVQHTMACHLVETRTIRAILFCQORDILLPNNNAVLVASGV	180	
Qy	181	ASNPIFIRRALLELTMAATQCTLLCPPRRLCTDNGIMIAMNGIERLRAGLGIHIDIEIRRE	240	
Db	181	ASNPIFIRRALLELTMAATQCTLLCPPRRLCTDNGIMIAMNGIERLRAGLGIHIDIEIRRE	240	
Qy	241	PKCPILGVDISKVEGEASIKVPOLKKEI	267	
Db	241	PKCPILGVDISKVEGEASIKVPOLKKEI	267	

RESULT 3
US-10-651-722-22
; Sequence 22, Application US/10651722
; Publication No. US20040048302A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1
; FILE REFERENCE: D0073 DIV
; CURRENT APPLICATION NUMBER: US/10/651,722
; CURRENT FILING DATE: 2003-08-29

```

? PRIOR APPLICATION NUMBER: US 60/266,518
? PRIOR FILING DATE: 2001-02-05
? PRIOR APPLICATION NUMBER: US 10/067,443
? PRIOR FILING DATE: 2002-02-05
? PRIOR APPLICATION NUMBER: US 60/282,814
? PRIOR FILING DATE: 2001-04-10
? NUMBER OF SEQ ID NOS: 71
? SOFTWARE: PatentIn version 3.2
? SEQ ID NO 22
? LENGTH: 267
? TYPE: prt
? ORGANISM: homo sapiens
? US-10-651-722-22

```

Query Match	100.0%	Score 1385	DB 15	Length 267
Best Local Similarity	100.0%	Pred. No. 2,2e-139		
Matches 267	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	1	MEAAHATLTRLTNKKEFPFLVLLISGGHCLLAVGVSDPFLILGKSLDIPADMDLKVARR	60	
DB	1	MEAAHATLTRLTNKKEFPFLVLLISGGHCLLAVGVSDPFLILGKSLDIPADMDLKVARR	60	
QY	61	LSLLIKHEPCSTMSGGKALIEHLAKQGNRPHDIKPELHHAKKCDPSFTGLQHVTDKIMKK	120	
DB	61	LSLLIKHEPCSTMSGGKALIEHLAKQGNRPHDIKPELHHAKKCDPSFTGLQHVTDKIMKK	120	
QY	121	EKEEGIEKQGLSSAADIAATVQHTMACHLYKTRHRAILFKQKRDLPQNNAAVLVAASGV	180	
DB	121	EKEEGIEKQGLSSAADIAATVQHTMACHLYKTRHRAILFKQKRDLPQNNAAVLVAASGV	180	
QY	181	ASNEYIRALAEILTNATQCTLLCPPRLCTNGMIANNGIERLPAAGIGLIHIDIEGIRYE	240	
DB	181	ASNEYIRALAEILTNATQCTLLCPPRLCTNGMIANNGIERLPAAGIGLIHIDIEGIRYE	240	
QY	241	PKCPLGVDISKVEGEASIKVPOLKKEI	267	
DB	241	PKCPLGVDISKVEGEASIKVPOLKKEI	267	

```

RESULT 4
US-10-067-443-2
; Sequence 2, Application US/10067443
; Publication No. US20030082782A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING
; TITLE OF INVENTION: SPINAL CORD, MP-1
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 414
; TYPE: prt
; ORGANISM: Homo sapiens
US-10-067-443-2

```

	Query Match	100.0%	Score 1385;	DB 14;	Length 414;
	Best Local Similarity	100.0%	Pred. NO. 4.2e-139;		
	Matches	267;	Conservative	0;	Mismatches 0; Indels 0; Gaps 0
QY	1	MEAAHATRLTNKVFPEFLVLLISGGCHLALVQGVSDFLIKGSLDIAPGMDLKVARR	60		
DB	148	MEAAHATRLTNKVFPEFLVLLISGGCHLALVQGVSDFLIKGSLDIAPGMDLKVARR	207		
QY	61	LSLIHQPECSIMSGGKALEHLAKQGNRFPHFDIKPELIHAKKNCDFSTGQHVTYDKIIMKK	120		
DB	208	LSLIHQPECSIMSGGKALEHLAKQGNRFPHFDIKPELIHAKKNCDFSTGQHVTYDKIIMKK	267		

```
QY 121 EKEGIEKGQILSSAADIATVQHTMACHLVKTRTHRALIFCKORDLLPQNNAVLVASGV 180
| | | | |
DB 268 EKEGIEKGQILSSAADIATVQHTMACHLVKTRTHRALIFCKORDLLPQNNAVLVASGV 327
QY 181 ASNFYIRRAELITNATQCTLLCPPRLCTDNGMIAMNGIERLRAGLGILHDIEGIRYE 240
| | | | |
DB 328 ASNFYIRRAELITNATQCTLLCPPRLCTDNGMIAMNGIERLRAGLGILHDIEGIRYE 387
QY 241 PKCPLGVDISKVEGASIKVPOLKMEI 267
| | | | |
DB 388 PKCPLGVDISKVEGASIKVPOLKMEI 414
```

```
RESULT 5
US-10-649-273-2
; Sequence 2, Application US/10649273
; Publication No. US20040043407A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1
; FILE REFERENCE: D0073 CNT
; CURRENT APPLICATION NUMBER: US/10/649, 273
; PRIOR FILING DATE: 2003-08-27
; PRIOR APPLICATION NUMBER: US 60/266, 518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067, 443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282, 814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-649-273-2
```

```
Query Match 100.0%; Score 1385; DB 15; Length 414;
Best Local Similarity 100.0%; Pred. No. 4.2e-139;
Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 MEAHALTIRLTNKVEPPLVLLISGHCILALVQGVSDPFLIGKSLDIAPGMDLKVARR 60
| | | | |
DB 148 MEAHALTIRLTNKVEPPLVLLISGHCILALVQGVSDPFLIGKSLDIAPGMDLKVARR 207
QY 61 LSLIKHPECSTWGGKALIEHLAKQGNRPHFDIKPPLHAKNCDPSTFTGLQHTVDKIMKK 120
| | | | |
DB 208 LSLIKHPECSTWGGKALIEHLAKQGNRPHFDIKPPLHAKNCDPSTFTGLQHTVDKIMKK 267
QY 121 EKEGIEKGQILSSAADIATVQHTMACHLVKTRTHRALIFCKORDLLPQNNAVLVASGV 180
| | | | |
DB 268 EKEGIEKGQILSSAADIATVQHTMACHLVKTRTHRALIFCKORDLLPQNNAVLVASGV 327
QY 181 ASNFYIRRAELITNATQCTLLCPPRLCTDNGMIAMNGIERLRAGLGILHDIEGIRYE 240
| | | | |
DB 328 ASNFYIRRAELITNATQCTLLCPPRLCTDNGMIAMNGIERLRAGLGILHDIEGIRYE 387
QY 241 PKCPLGVDISKVEGASIKVPOLKMEI 267
| | | | |
DB 388 PKCPLGVDISKVEGASIKVPOLKMEI 414
```

```
RESULT 6
US-10-651-722-2
; Sequence 2, Application US/10651722
; Publication No. US20040048302A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1
; FILE REFERENCE: D0073 DIV
; CURRENT APPLICATION NUMBER: US/10/651, 722
; PRIOR FILING DATE: 2003-08-29
```

```
; PRIOR APPLICATION NUMBER: US 60/266, 518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067, 443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282, 814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-651-722-2
```

```
Query Match 100.0%; Score 1385; DB 15; Length 414;
Best Local Similarity 100.0%; Pred. No. 4.2e-139;
Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEAHALTIRLTNKVEPPLVLLISGHCILALVQGVSDPFLIGKSLDIAPGMDLKVARR 60
| | | | |
DB 148 MEAHALTIRLTNKVEPPLVLLISGHCILALVQGVSDPFLIGKSLDIAPGMDLKVARR 207
QY 61 LSLIKHPECSTWGGKALIEHLAKQGNRPHFDIKPPLHAKNCDPSTFTGLQHTVDKIMKK 120
| | | | |
DB 208 LSLIKHPECSTWGGKALIEHLAKQGNRPHFDIKPPLHAKNCDPSTFTGLQHTVDKIMKK 267
QY 121 EKEGIEKGQILSSAADIATVQHTMACHLVKTRTHRALIFCKORDLLPQNNAVLVASGV 180
| | | | |
DB 268 EKEGIEKGQILSSAADIATVQHTMACHLVKTRTHRALIFCKORDLLPQNNAVLVASGV 327
QY 181 ASNFYIRRAELITNATQCTLLCPPRLCTDNGMIAMNGIERLRAGLGILHDIEGIRYE 240
| | | | |
DB 328 ASNFYIRRAELITNATQCTLLCPPRLCTDNGMIAMNGIERLRAGLGILHDIEGIRYE 387
QY 241 PKCPLGVDISKVEGASIKVPOLKMEI 267
| | | | |
DB 388 PKCPLGVDISKVEGASIKVPOLKMEI 414
```

```
RESULT 7
US-10-480-988-8
; Sequence 8, Application US/10480988
; Publication No. US20050069877A1
; GENERAL INFORMATION:
; APPLICANT: GANDHI, Ameena R.; KABLE, Amy B.;
; APPLICANT: SMARNAKAR, Anita; HAPALIA, April J.A.;
; APPLICANT: TRAN, Bao; DUGGAN, Brendan M.;
; APPLICANT: WARREN, Bridget A.; ISON, Craig H.;
; APPLICANT: HONCHELL, Cynthia D.; NGUYEN, Daniel B.;
; APPLICANT: LU, Dzung Aina M.; LEE, Ernestine A.;
; APPLICANT: YUE, Henry; FORSYTHE, Ian J.;
; APPLICANT: BARROSO, Ines; RAMKUMAR, Jayalakshi;
; APPLICANT: GRIFPIN, Jennifer A.; LI, Joana X.;
; APPLICANT: YANG, Junming; THANGAVELU, Kavitha;
; APPLICANT: GIERZEN, Kimberly J.; DING, Li;
; APPLICANT: BAUGHN, Mariah R.; BOROWSKY, Mark L.;
; APPLICANT: YAO, Monique G.; CHAMLA, Nandinder K.;
; APPLICANT: MASON, Patricia M.; GURURAJAN, Rajagopal;
; APPLICANT: LEE, Sally; BECHA, Shanya D.;
; APPLICANT: LEE, Soo Yeun; TRAN, Uyen K.;
; APPLICANT: ELIOTT, Vicki S.; LUO, Men;
; APPLICANT: SPRAGUE, William W.; TANG, Y. Tom;
; APPLICANT: LU, Yan; ZEBARADIAN, Yeganeh
; TITLE OF INVENTION: PROTEIN MODIFICATION AND MAINTENANCE MOLECULES
; FILE REFERENCE: PF-1040 USN
; CURRENT APPLICATION NUMBER: US/10/480, 988
; PRIOR FILING DATE: 2003-12-16
; PRIOR APPLICATION NUMBER: PCT/US02/19360
; PRIOR FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: US 60/300, 508
; PRIOR FILING DATE: 2001-06-22
; PRIOR APPLICATION NUMBER: US 60/303, 445
; PRIOR FILING DATE: 2001-07-06
```

;; PRIOR APPLICATION NUMBER: US 60/305,405
;; PRIOR FILING DATE: 2001-07-13
;; PRIOR APPLICATION NUMBER: US 60/311,442
;; PRIOR FILING DATE: 2001-08-09
;; PRIOR APPLICATION NUMBER: US 60/314,821
;; PRIOR FILING DATE: 2001-08-24
;; PRIOR APPLICATION NUMBER: US 60/315,992
;; PRIOR FILING DATE: 2001-08-29
;; PRIOR APPLICATION NUMBER: US 60/378,205
;; PRIOR FILING DATE: 2002-05-03
;; NUMBER OF SEQ ID NOS: 56
;; SOFTWARE: PERL Program
;; SEQ ID NO 8
;; LENGTH: 414
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; FEATURE: misc_feature
;; NAME/KEY: misc_feature
;; OTHER INFORMATION: Incyte ID No: 7632424CD1
US-10-480-988-8

Query Match 100.0%; Score 1385; DB 17; Length 414;
Best Local Similarity 100.0%; Pred. No. 4.2e-139;
Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MEAHALTIRLTNKVEPPLVLLISGGHCLALVOGVSDPFLLLGKSLDIAPGMDLVARR 60
Db 148 MEAHALTIRLTNKVEPPLVLLISGGHCLALVOGVSDPFLLLGKSLDIAPGMDLVARR 207
Qy 61 LSLIKHECSTMSGKAI EHLAKQGNRFHDIKPLHAKNCDPFTGLOHTVDKIIIMKK 120
Db 208 LSLIKHECSTMSGKAI EHLAKQGNRFHDIKPLHAKNCDPFTGLOHTVDKIIIMKK 267
Qy 121 EKEGIRKQGLSSAADIAATVOHTMACHLVKRTIRALLFCQORLLPQNNAVLVASGV 180
Db 268 EKEGIRKQGLSSAADIAATVOHTMACHLVKRTIRALLFCQORLLPQNNAVLVASGV 327
Qy 181 ASNFYIRALTEILTNATQCTLLCPPRLCTDNGIMIANNGIERLHAGILHDIIGIRYE 240
Db 328 ASNFYIRALTEILTNATQCTLLCPPRLCTDNGIMIANNGIERLHAGILHDIIGIRYE 387
Qy 241 PKCPLGVDISKVEGASIKVQLKMEI 267
Db 388 PKCPLGVDISKVEGASIKVQLKMEI 414

RESULT 8
US-10-067-443-19
;; Sequence 19, Application US/10067443
;; Publication No. US20030082782A1
;; GENERAL INFORMATION:
;; APPLICANT: Bristol-Myers Squibb Company
;; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED
;; FILE REFERENCE: D0073 NP
;; CURRENT APPLICATION NUMBER: US/10/067,443
;; CURRENT FILING DATE: 2002-02-05
;; PRIOR APPLICATION NUMBER: US 60/266,518
;; PRIOR FILING DATE: 2001-02-05
;; PRIOR APPLICATION NUMBER: US 60/282,814
;; PRIOR FILING DATE: 2001-04-10
;; NUMBER OF SEQ ID NOS: 71
;; SOFTWARE: PatentIn version 3.0
;; SEQ ID NO 19
;; LENGTH: 439
;; TYPE: PRT
;; ORGANISM: homo sapiens
US-10-067-443-19

Query Match 98.1%; Score 1358.5; DB 14; Length 439;
Best Local Similarity 91.1%; Pred. No. 3.2e-136;
Matches 266; Conservative 0; Mismatches 1; Indels 25; Gaps 1;

Qy 1 MEAHALTIRLTNKVEPPLVLLISGGHCLALVOGVSDPFLLLGKSLDIAPGMDLVARR 60
Db 148 MEAHALTIRLTNKVEPPLVLLISGGHCLALVOGVSDPFLLLGKSLDIAPGMDLVARR 207
Qy 61 LSLIKHECSTMSGKAI EHLAKQGNRFHDIKPLHAKNCDPFTGLOHTVDKIIIMKK 120
Db 208 LSLIKHECSTMSGKAI EHLAKQGNRFHDIKPLHAKNCDPFTGLOHTVDKIIIMKK 267
Qy 121 EKEGIRKQGLSSAADIAATVOHTMACHLVKRTIRALLFCQORLLPQNNAVLVASGV 180
Db 268 EKEGIRKQGLSSAADIAATVOHTMACHLVKRTIRALLFCQORLLPQNNAVLVASGV 327
Qy 181 ASNFYIRALTEILTNATQCTLLCPPRLCTDNGIMIANNGIERLHAGILHDIIGIRYE 240
Db 328 ASNFYIRALTEILTNATQCTLLCPPRLCTDNGIMIANNGIERLHAGILHDIIGIRYE 387
Qy 241 PKCPLGVDISKVEGASIKVQLKMEI 267
Db 388 PKCPLGVDISKVEGASIKVQLKMEI 439

RESULT 9
US-10-649-273-19
;; Sequence 19, Application US/10649273
;; Publication No. US20040043407A1
;; GENERAL INFORMATION:
;; APPLICANT: Bristol-Myers Squibb Company
;; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1
;; FILE REFERENCE: D0073 CNT
;; CURRENT APPLICATION NUMBER: US/10/649,273
;; CURRENT FILING DATE: 2003-08-27
;; PRIOR APPLICATION NUMBER: US 60/266,518
;; PRIOR FILING DATE: 2001-02-05
;; PRIOR APPLICATION NUMBER: US 10/067,443
;; PRIOR FILING DATE: 2002-02-05
;; PRIOR APPLICATION NUMBER: US 60/282,814
;; PRIOR FILING DATE: 2001-04-10
;; NUMBER OF SEQ ID NOS: 71
;; SOFTWARE: PatentIn version 3.2
;; SEQ ID NO 19
;; LENGTH: 439
;; TYPE: PRT
;; ORGANISM: homo sapiens
US-10-649-273-19

Query Match 98.1%; Score 1358.5; DB 15; Length 439;
Best Local Similarity 91.1%; Pred. No. 3.2e-136;
Matches 266; Conservative 0; Mismatches 1; Indels 25; Gaps 1;

Qy 1 MEAHALTIRLTNKVEPPLVLLISGGHCLALVOGVSDPFLLLGKSLDIAPGMDLVARR 60
Db 148 MEAHALTIRLTNKVEPPLVLLISGGHCLALVOGVSDPFLLLGKSLDIAPGMDLVARR 207
Qy 61 LSLIKHECSTMSGKAI EHLAKQGNRFHDIKPLHAKNCDPFTGLOHTVDKIIIMKK 120
Db 208 LSLIKHECSTMSGKAI EHLAKQGNRFHDIKPLHAKNCDPFTGLOHTVDKIIIMKK 267
Qy 121 EKEGIRKQGLSSAADIAATVOHTMACHLVKRTIRALLFCQORLLPQNNAVLVASGV 180
Db 268 EKEGIRKQGLSSAADIAATVOHTMACHLVKRTIRALLFCQORLLPQNNAVLVASGV 327
Qy 181 ASNFYIRALTEILTNATQCTLLCPPRLCTDNGIMIANNGIERLHAGILHDIIGIRYE 240
Db 328 ASNFYIRALTEILTNATQCTLLCPPRLCTDNGIMIANNGIERLHAGILHDIIGIRYE 387
Qy 241 PKCPLGVDISKVEGASIKVQLKMEI 267
Db 388 PKCPLGVDISKVEGASIKVQLKMEI 439

RESULT 10
US-10-651-722-19
;; Sequence 19, Application US/10651722

```
Publication No. US20040048302A1
GENERAL INFORMATION:
APPLICANT: Bristol-Myers Squibb Company
TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEINASE, MP-1
FILE REFERENCE: 00073 DIV
CURRENT APPLICATION NUMBER: US/10/651,722
PRIOR FILING DATE: 2003-08-29
PRIOR APPLICATION NUMBER: US 60/266,518
PRIOR FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: US 10/067,443
PRIOR FILING DATE: 2002-02-05
PRIOR APPLICATION NUMBER: US 60/282,814
PRIOR FILING DATE: 2001-04-10
NUMBER OF SEQ ID NOS: 71
SOFTWARE: PatentIn version 3.2
SEQ ID NO 19
LENGTH: 439
TYPE: PRT
ORGANISM: homo sapiens
US-10-651-722-19

Query Match          98.1%; Score 1358.5; DB 15; Length 439;
Best Local Similarity 91.1%; Pred. No. 3.2e-136;
Matches 266; Conservative 0; Mismatches 1; Indels 25; Gaps 1;

QY 1 MEAHLTIRLTNKKVEFPFLVLLISGHCCLALVOGVSDFLLGKSLDIAPGMDLKVARR 60
DB 148 MEAHLTIRLTNKKVEFPFLVLLISGHCCLALVOGVSDFLLGKSLDIAPGMDLKVARR 207
QY 61 LSLIKHPECSTMSGGKAI EHLAKQGNRFHDIKPLHAKNCDFSFTGLQHTYDVKIMKK 120
DB 208 LSLIKHPECSTMSGGKAI EHLAKQGNRFHDIKPLHAKNCDFSFTGLQHTYDVKIMKK 267
QY 121 EKEEGI-----EKGQIISADIAATVQHTMACHLVKRTK 155
DB 268 EKEEGIFLISKVEQINIGLCLKIAHFCRYEKQIISADIAATVQHTMACHLVKRTK 327
QY 156 RALIFCQKQDLTPONNAVTVASGVASNFYIRALLETITNAOCTLCPPRLCTDNGIM 215
DB 328 RALIFCQKQDLTPONNAVTVASGVASNFYIRALLETITNAOCTLCPPRLCTDNGIM 387
QY 216 IAWNGIERLRAGLGLHDIEGIRYEPKCPGLGVDSKEVGEASIKVPOLKMEI 267
DB 388 IAWNGIERLRAGLGLHDIEGIRYEPKCPGLGVDSKEVGEASIKVPOLKMEI 439

RESULT 11
US-10-012-140-5
Sequence 5, Application US/10012140
Publication No. US20030009017A1
GENERAL INFORMATION:
APPLICANT: Leibdy, Kevin R.
APPLICANT: Kapeller-Libermann, Rosana
APPLICANT: Gluckemann, Maria A.
TITLE OF INVENTION: 38650, 28472, 5495, 65507, 81588, AND
TITLE OF INVENTION: 14354 METHODS AND COMPOSITIONS OF HUMAN PROTEINS AND USES
TITLE OF INVENTION: THEREOF
FILE REFERENCE: 381552004900
CURRENT APPLICATION NUMBER: US/10/012,140
PRIOR FILING DATE: 2001-11-08
PRIOR APPLICATION NUMBER: 60/246,768
PRIOR FILING DATE: 2000-11-08
PRIOR APPLICATION NUMBER: 60/246,772
PRIOR FILING DATE: 2000-11-08
PRIOR APPLICATION NUMBER: 60/249,185
PRIOR FILING DATE: 2000-11-15
NUMBER OF SEQ ID NOS: 49
SOFTWARE: PatSeq for Windows Version 4.0
SEQ ID NO 5
LENGTH: 414
TYPE: PRT
ORGANISM: Homo sapiens
US-10-012-140-5
```

```
Query Match          97.3%; Score 1348; DB 14; Length 414;
Best Local Similarity 97.4%; Pred. No. 3.9e-135;
Matches 260; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 MEAHLTIRLTNKKVEFPFLVLLISGHCCLALVOGVSDFLLGKSLDIAPGMDLKVARR 60
DB 148 MEAHLTIRLTNKKVEFPFLVLLISGHCCLALVOGVSDFLLGKSLDIAPGMDLKVARR 207
QY 61 LSLIKHPECSTMSGGKAI EHLAKQGNRFHDIKPLHAKNCDFSFTGLQHTYDVKIMKK 120
DB 208 LSLIKHPECSTMSGGKAI EHLAKQGNRFHDIKPLHAKNCDFSFTGLQHTYDVKIMKK 267
QY 121 EKEEGI EKGQIISADIAATVQHTMACHLVKRTKRALIFCQKQDLTPONNAVTVASGV 180
DB 268 EKEEGIEKGQIISADIAATVQHTMACHLVKRTKRALIFCQKQDLTPONNAVTVASGV 327
QY 181 ASNFYIRALLETITNAOCTLCPPRLCTDNGIMIAWNGIERLRAGLGLHDIEGIRY 240
DB 328 ASNFYIRALLETITNAOCTLCPPRLCTDNGIMIAWNGIERLRAGLGLHDIEGIRY 387
QY 241 PKCPLGVDSKEVGEASIKVPOLKMEI 267
DB 388 PKCPLGVDSKEVGEASIKVPOLKMEI 414
```

```
RESULT 12
US-10-094-749-2039
Sequence 2039, Application US/10094749
Publication No. US20030219741A1
GENERAL INFORMATION:
APPLICANT: ISOGAI, TAKAO
APPLICANT: SUGIYAMA, TOMOYASU
APPLICANT: OTSUKI, TETSUJI
APPLICANT: WAKAMATSU, AI
APPLICANT: SATO, HIROYUKI
APPLICANT: ISHII, SHIZUKO
APPLICANT: YAMAMOTO, JUN-ICHI
APPLICANT: ISONO, YUUKO
APPLICANT: HIO, YURI
APPLICANT: OTSUKA, KAORI
APPLICANT: NAGAI, KEIICHI
APPLICANT: IRIE, RYOTARO
APPLICANT: TAMECHIKA, ICHIRO
APPLICANT: SEKI, NAOHITO
APPLICANT: YOSHIKAWA, TSUTOMU
APPLICANT: OTSUKA, MOTOYUKI
APPLICANT: NAGAHARI, KENJI
APPLICANT: MASUHO, YASUHIKO
TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
FILE REFERENCE: 084335/0160
CURRENT APPLICATION NUMBER: US/10/094,749
PRIOR FILING DATE: 2002-03-12
PRIOR APPLICATION NUMBER: 60/350,435
PRIOR FILING DATE: 2002-01-24
PRIOR APPLICATION NUMBER: JP 2001-328381
PRIOR FILING DATE: 2001-09-14
NUMBER OF SEQ ID NOS: 3381
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2039
LENGTH: 364
TYPE: PRT
ORGANISM: Homo sapiens
US-10-094-749-2039

Query Match          79.8%; Score 1105; DB 15; Length 364;
Best Local Similarity 98.6%; Pred. No. 3.4e-109;
Matches 214; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MEAHLTIRLTNKKVEFPFLVLLISGHCCLALVOGVSDFLLGKSLDIAPGMDLKVARR 60
DB 148 MEAHLTIRLTNKKVEFPFLVLLISGHCCLALVOGVSDFLLGKSLDIAPGMDLKVARR 207
```

```
QY 61 LSLIKPECGTSGGKAI EHLAKQGNFHDIKPPLHAKNCDSFTGLQHVTDKIMKK 120
DB 208 LPLIKPECGTSGGKAI EHLAKQGNFHDIKPPLHAKNCDSFTGLQHVTDKIMKK 267
QY 121 EKEEGIEKGQILSSAADIAATVQHTMACHLVKRTFRALLPKQKRDLLPONNAVIVASGV 180
DB 268 EKEEGIEKGQILSSAADIAATVQHTMACHLVKRTFRALLPKQKRDLLPONNAVIVASGV 327
QY 181 ASNFYIRRALEILTNAQTCTLLCPPRLCTDNGIMIA 217
DB 328 ASNFYIRRALEILTNAQTCTLLCPPRLCTDNGIMIA 364

RESULT 13
US-10-424-599-209259
; Sequence 209259, Application US/10424599
; Publication NO. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 209259
; LENGTH: 445
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_3098C.1.pep
US-10-424-599-209259

Query Match 30.5%; Score 422.5; DB 15; Length 445;
Best Local Similarity 41.6%; Pred. No. 5.4e-36;
Matches 104; Conservative 32; Mismatches 89; Indels 25; Gaps 6;

QY 1 MEAHLTIRLTNK-VEPPFLVLLISGGHCLALVGVSDFLILGSLDIAPGDMLDKVAR 59
DB 165 MEAHLVGRLEIKDLPFPFALLISGGHNLVLARDLQYIQLGTTIDDAIGEAYDKTAK 224
QY 60 RLSLIKPECGTSGGKAI EHLAKQGNFHDIKPPLHAKNCDSFTGLQHVTDKIMKK 114
DB 225 WLGLDLR-----RGGPAIEKLAEGNAESVKSIPMKOHKDCNFSYAGLKTQVRLAES 279
QY 115 KIIMKEKEEGIEKGQILSSAADIAATVQHTMACHLVKRTFRALLPKQKRDLLPONNAVIL 174
DB 280 KIDAKIPISASNGDRL-SPADIASFORIAVHLHERCERAIQWALKEPSTIRH---L 335
QY 175 VASGVASNFYIRRALEILTNAQTCTLLCPPRLCTDNGIMIAMNGIERLPAAGLILHDI 234
DB 336 VVSGVASNQVRRALDVMVKKNGQLVCPPLRLCTDNGVMIAMGIEHFRMG----- 388
QY 235 EGIRYEPKCP 244
DB 389 ---RYDPPPP 395

RESULT 14
US-11-097-143-34191
; Sequence 34191, Application US/11097143
; Publication No. US2005020858A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; FILE REFERENCE: CLO00728
; CURRENT APPLICATION NUMBER: US/11/097,143
```

```
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34191
; LENGTH: 409
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-11-097-143-34191

Query Match 30.0%; Score 415.5; DB 20; Length 409;
Best Local Similarity 38.2%; Pred. No. 2.7e-35;
Matches 100; Conservative 43; Mismatches 104; Indels 15; Gaps 6;

QY 1 MEAHLTIRLTNK-KVEPPFLVLLISGGHCLALVGVSDFLILGSLDIAPGDMLDKVAR 58
DB 136 MEAHLQARMEHPEDIGYFFLCLLASGCHOLVANGPRLTLGTLDDAPGEAFDCKII 195
QY 59 RLSLIKPECGTSGGKAI EHLAKQGNFHDIKPPLHAKNCDSFTGLQHVTDKIMKK 117
DB 196 RLRLHIILPEYRLMNGGRALIEHAQASDPLAIEPPLAQORNCNFSFAGIKNNSFRAL 255
QY 118 MKKEKEEGIEKGQILSSAADIAATVQHTMACHLVKRTFRALLPKQKRDLLPONNAVIL 175
DB 256 RARERAEPTPPDGVISNYDPCAGILRSVSRHLMHRTQRAIEYCLLPHQLPEDTPTLY 315
QY 176 ASGVASNFYIRRALEILTNAQTCTLLCPPRLCTDNGIMIAMNGIERLPAAGLILHDI 235
DB 316 MSGVANNDAIVANIEHLAAQVGRSFRPSKRYCSNNGVMIAMHGYEQL-----LQDXE 369
QY 236 -GIRYEPKCPGLVDISKVGEA 256
DB 370 ASTRYDYD---SIDIQGSAGFA 388

RESULT 15
US-10-067-443-3
; Sequence 3, Application US/10067443
; Publication No. US20030082782A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE HIGHLY EXPRESSED IN
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 463
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-067-443-3
```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2005, 15:23:22 ; Search time 16.5424 Seconds
(without alignments)
1552.972 Million cell updates/sec

Title: US-10-649-273-2_COPY_148_414

Perfect score: 1385

Sequence: 1 MEAHALTIRLTNKVPEPFLV.....DISKVEBASIKVPOLKMEI 267

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: -pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	409.5	29.6	463	2	probable O-sialogl
2	338	24.4	463	2	O-sialoglycoprotei
3	338	24.4	366	2	probable O-sialogl
4	338	24.4	387	2	probable O-sialogl
5	334.5	24.2	359	2	O-sialoglycoprotei
6	329.5	23.8	344	2	O-sialoglycoprotei
7	326.5	23.6	367	2	peptidase M22 fami
8	291.5	21.0	341	2	O-sialoglycoprotei
9	290.5	21.0	342	2	O-sialoglycoprotei
10	289.5	20.9	342	2	O-sialoglycoprotei
11	279.5	20.2	337	2	probable glycoprot
12	279.5	20.2	337	2	probable glycoprot
13	278.5	20.1	337	2	probable O-sialogl
14	278.5	20.1	337	2	probable O-sialogl
15	276	19.9	421	2	hypothetical prote
16	275.5	19.9	337	2	O-sialoglycoprotei
17	275	19.9	412	2	probable proteinas
18	270.5	19.5	335	2	O-sialoglycoprotei
19	265.5	19.2	354	2	O-sialoglycoprotei
20	263.5	19.0	354	2	probable O-sialogl
21	263	18.5	346	2	O-sialoglycoprotei
22	256	18.5	346	2	O-sialoglycoprotei
23	253	18.3	336	2	O-sialoglycoprotei
24	250.5	18.1	346	2	glycoprotein endop
25	250	18.1	336	2	glycoproteinase fa
26	244.5	17.7	343	2	glycoprotein endop
27	243	17.5	348	2	hypothetical prote
28	242.5	17.4	346	2	O-sialoglycoprotei
29	241.5	17.4	346	2	O-sialoglycoprotei

30	239.5	17.3	344	2	AC1334	glycoprotein endop
31	237.5	17.1	340	2	B97011	probable O-sialogl
32	236.5	17.1	336	2	E84936	O-sialoglycoprotei
33	233	16.8	323	2	G69388	O-sialoglycoprotei
34	231	16.7	344	2	AB1705	glycoprotein endop
35	229.5	16.6	346	2	G86661	O-sialoglycoprotei
36	222.5	16.1	324	2	F75029	O-sialoglycoprotei
37	220	15.9	338	2	A71545	probable O-sialogl
38	219.5	15.8	534	2	H69056	O-sialoglycoprotei
39	219	15.8	335	2	E81278	probable glycoprot
40	218	15.7	344	2	H72106	O-sialoglycoprotei
41	218	15.7	344	2	B86515	O-sialoglycoprotei
42	217	15.7	341	2	G89996	hypothetical prote
43	216	15.6	407	2	S50740	QR17 protein - yea
44	215.5	15.6	324	2	C71215	O-sialoglycoprotei
45	215.5	15.6	344	2	H70737	probable O-sialogl

ALIGNMENTS

```
RESULT 1
E84888
Probable O-sialoglycoprotein endopeptidase [imported] - Arabidopsis thaliana
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Jun-2003
C/Accession: E84888
R.lin. X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Taiton, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Frazer, C.M.; Venter, J.
Nature 402, 761-768, 1999
A/Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A/Reference number: A64420; MUID:20083487; PMID:10617197
A/Accession: E84888
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-463 <STO>
A/Cross-references: GB:AE002093; NID:g2583127; PIDN:AAB82636.1; GSPDB:GN00139
C/Genetics:
A/Gene: At2g45270
A/Map position: 2
C/Superfamily: O-sialoglycoprotein endopeptidase

Query Match          29.6% Score 409.5; DB 2; Length 463;
Best Local Similarity 36.6%; Pred. No. 5.1e-29;
Matches 100; Conservative 37; Mismatches 91; Indels 45; Gaps 6;

QY      1 MEAHALTIRLT-NKVEPFLVLLISGHCITLAVGVSDPFLIGKSLDAPGMDLKVAR 59
        |||||  ||  :: |||||  |||||  ||  :  :: |||||  |||||
DB      192 MEAHALTIRLVGQELSPFPMALLISGHNLLVLAHKLGQYTGTVDDAIGAPDKTAK 251
        |||||  ||  :: |||||  |||||  ||  :  :: |||||  |||||
QY      60 RLSTLRKPECSMGSGKAIENLAKQGNRFHPIKRPPLHAKNCDFSTGLQVHTDKIKK 119
        |||||  ||  :: |||||  |||||  ||  :  :: |||||  |||||
DB      252 WLGIDMH-----RSGCPVBEILTEGDAKSVFNNPMKHKCNFSAIGLKTQVRLAIA 306
        |||||  ||  :: |||||  |||||  ||  :  :: |||||  |||||
QY      120 KEKEGIRKQGLTSSAADIATVQHTMACHVAKTRHATLFCQKORLLQNNANVLVASGG 179
        |||||  ||  :: |||||  |||||  ||  :  :: |||||  |||||
DB      307 KE-----IRRADPAASFQVAVLHLEKCKERALDMLR--LEPSIKHMTISGG 353
        |||||  ||  :: |||||  |||||  ||  :  :: |||||  |||||
QY      180 VASNFYIRRALEILTNAQCTLLCPPRPRLCTDNGIMIANNGIERLHAGIHLHDEGIRY 239
        |||||  ||  :: |||||  |||||  ||  :  :: |||||  |||||
DB      354 VASNKYVLRLLNINIVENKNLKVCPPLSLCTDNGVAVMTGLEHFRVGS-----RY 403
        |||||  ||  :: |||||  |||||  ||  :  :: |||||  |||||
QY      240 E-----KCPPIGVSDISKVEGSA 256
        :  |||  :  |||  :  |||  :  |||  :  |||  :  |||  :  |||
DB      404 DPPPATPEPDYVDLRPRMPLGEERYAKGRSEA 436
        :  |||  :  |||  :  |||  :  |||  :  |||  :  |||  :  |||

RESULT 2
AB2902
O-sialoglycoprotein endopeptidase gcp [imported] - Agrobacterium tumefaciens (strain C58
C/Species: Agrobacterium tumefaciens
C/Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
```


C;Accession: AB2902
R:Wood, D.W.; Seibubel, J.C.; Kaul, R.; Monks, C.D.; Chen, L.; Wood, G.B.; Chen, Y.; Woo, I.
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McCell
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, X.; Biddle, P.; Jung, M.; Krepsan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.

A>Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577, PMID:21608550, PMID:11743193
A:Accession: AB2902
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-365 <KUR>
A:Cross-references: GB:AE008688; PIDN:AAL43632.1; PID:g17741154; GSPDB:GN00186
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: gcp
A:Map position: circular chromosome
C:Superfamily: O-sialoglycoprotein endopeptidase

Query Match 24.4%; Score 338; DB 2; Length 365;
Best Local Similarity 34.9%; Pred. No. 1.1e-22;
Matches 90; Conservative 40; Mismatches 98; Indels 30; Gaps 7;

OY 1 MEAHATITLTKNVPEPFLVLLISGGHCLTALVGVSDFLLGKSIDIAFGDMLDKVAR 60
| : | | | : | : | : | : | : | : | : | : | : | : | : | : | :
Db LEGHALTARLTDTGISFYLMTLVSGHTQLVALVRVGEYERWGTIIDDLAGEAFDTAKL 179
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
OY 61 LSILIKPECSWTMSGKAIEHLAKOGNPFPDIKPRLHAKNCDFSTGL----QHYTDKI 116
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 180 LGI-PYP-----GRAVENAAAGDPDRFPLPRMWGEARLDIFSGLKTAVROATAI 232
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
OY 117 IMKEEKGIEKKQIISSAADIAATVOHTMACHLVKTRTHAILFCOKRDILLPONNA--VL 174
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 233 APISEOD-----IADICASFQKAVASRTLKDRIGRGAREFYV--PPHINGEPAL 279
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
OY 175 VASGVASNRYIRALEILTNAOTCTLLCPFRCTDNCGIMAMNGINIELRAGLGILHI 234
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 280 VVAGGAANOETRIQTALCDTHGFHFVAAPHRLCTDNAAMTAMAGLERMAG---RQA 335
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
OY 235 EGIRVEPKCELPDISKE 252
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 336 DALEVAAPRSRWPIDGSAE 353
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

RESULT 3
D97677
Probable o-sialoglycoprotein endopeptidase (glycoproteinnase) [imported] - Agrobacterium
C.Species: Agrobacterium tumefaciens
C.Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 18-Nov-2002
C.Accession: D97677
R:Goodner, F.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wolam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.,
Science 294, 2323-2328, 2001
A>Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tun
A:Reference number: AY9359; PMID:21608551; PMID:11743194
A:Accession: D97677
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-366 <KUR>
A:Cross-references: GB:AE007869; PIDN:AAK8373.1; PID:g15157858; GSPDB:GN00169
C:Genetics:
A:Gene: AGR_C 4806
A:Map position: circular chromosome
C:Superfamily: O-sialoglycoprotein endopeptidase

Query Match 24.4%; Score 338; DB 2; Length 366;
Best Local Similarity 34.9%; Pred. No. 1.1e-22;
Matches 90; Conservative 40; Mismatches 98; Indels 30; Gaps 7;

OY 1 MEAHATITLTKNVPEPFLVLLISGGHCLTALVGVSDFLLGKSIDIAFGDMLDKVAR 60
| : | | | | : | : | : | : | : | : | : | : | : | : | : | : |
Db LEGHALTARLTDTGISFYLMTLVSGHTQLVALVRVGEYERWGTIIDDLAGEAFDTAKL 180
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

Oy 1SLIHPEECSTAGGCAIETHLAKQGNRFHFDIKPELTHAKNCDSFSTGVL-----QVATDKI 116
 Db 181 LGL-PYP-----GGPAAVMAAAKGDPRDRPPLRPFWGGEARLDPSFSGLKTAARQAAATPI 233
 Oy 117 IMKEKEGIEKGQILSSADIAATVQHTMACHLVKRTTHAILFCQKQDRLPQNNAA--VL 174
 Db 234 APLSBOD-----IADICASFQKAVSRFLTKDRIGRGLARPKVE--PPHINGEPAL 280
 Oy 175 VASGGVANSFYIRRLAEITLTNAQTCLCPRPRLCTDNGMIAMNGIERLAAGLGIHDI 234
 Db 281 VVAGGVANAOEIRQTALQALCDTHGFRFPVAPRRLCTDNNAMIAWAGLERMAEG----ROA 336
 Oy 235 EGIRYEPKCPGLGVDISKE 252
 Db 337 DALBVAAPRSRWPLDGSAE 354
 RESULT 4
 E71711
 Probable O-sialoglycoprotein endopeptidase (gcp) RP037 - Rickettsia prowazekii
 C:Species: Rickettsia prowazekii
 C:Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 09-Jul-2004
 C:Accession: E71711
 R:Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sichevitz-Ponten, T.; Alsmark, U.
 Nature 396, 133-140, 1998
 A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.
 A:Reference number: A71630; MUID:99039499; PMID:9823893
 A:Accession: E71711
 A:Status: Preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-387 <AND>
 A:Cross-references: UNIPROT:Q9ZEAB; GB:AJ235270; GB:AJ235269; NID:g3860572; PIDD:CAA1450
 A:Experimental source: strain Madrid B
 C:Genetics:
 A:Gene: gcp, RP037
 C:Superfamily: O-sialoglycoprotein endopeptidase
 Query Match 24.4%; Score 338; DB 2; Length 387;
 Best Local Similarity 32.5%; Pred. No. 1,1e-22;
 Matches 87; Conservative 38; Mismatches 83; Indels 60; Gaps 6;
 Oy 1 MEAAHTIRLTNKEPFPVLVLSGGHCLLALVQGVSDFLLGKSLDIAPGDMLDKVAR 60
 Db 113 LEGHALTATLTNISYPVLLLASGGHCQFVAVLGLKKYKILGTTIIDAVGETPDKVAK 172
 Oy 61 LSLIHPEECSTAGGCAIETHLAKQGNRFHFDIKPELTHAKNCDSFSTGVLQHTDKIMK- 119
 Db 173 LNL-----SPGGEIERRALQSNPKYKFPKPIINSGCNMSFSGLKTAVRTILNKL 225
 Oy 120 KEKEGIEKGQILSSADIAATVQHTMACHLVKRTTHAILFCQKQ----- 163
 Db 226 KEVNSV-----INDIAASFQITGALLSSMGQDAIRLYKQILNDIYEDINHPTKLN 277
 Oy 164 ----RD-----LDPON-----NAVLVASGGVANSFYIRRLAEILT 195
 Db 278 LKSPKDEBNMKPRLKCTPRKYRIHIONSYRSLNLDITVIAGGVAAANKYQIELSLDCTR 337
 Oy 196 ATOCTLLCPRPRLCTDNGMIAMNGIER 223
 Db 338 PYGRRLIAPPKMLCTDNNAMIAWAGLER 365
 RESULT 5
 AB3274
 O-sialoglycoprotein endopeptidase (EC 3.4.24.57) [imported] - Brucella melitensis (strain
 C:Species: Brucella melitensis
 C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 15-Feb-2002
 C:Accession: AB3274
 R:DeIvecchio, V.G.; Kaparatl, V.; Redkar, R.J.; Patra, G.; Mujter, C.; Los, T.; Ivanova,
 Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
 A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis

[illegible]

RESULT 7

peptidase M22 family protein [imported] - Caulobacter crescentus

C:Species: Caulobacter crescentus

C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001

C:Accession: F87257

R:Nickman, W.C.; Feldbylum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.B.; Laub, M.T.; Deboy, R.T.; Dodson, R.U.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon, n. J.; Ertolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M. Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A:Title: Complete Genome Sequence of Caulobacter crescentus.

A:Reference number: A87249; MUID:21173698; PMID:11259647

A:Accession: F87257

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-367 <STO>

A:Cross-references: GB:AE005673; NID:G13421168; PIDN:AAK22058.1; GSPDB:GN00148

C:Genetics:

A:Gene: CC0071

C:Superfamily: O-sialoglycoprotein endopeptidase

Query Match 23.6%; Score 326.5; DB 2; Length 367;

Best Local Similarity 36.3%; Pred. No. 1.2e-21;

Matches 89; Conservative 36; Mismatches 99; Indels 21; Gaps 6;

OY 1 MEAAHTITPLTKVKEFPFLVLLISGGHCLALVGVSDFLILGKSLDIAFGMDLKVARR 60

DB 124 LEGHAVSARLQADIAVPPFLILLVSGGHQQLLEVSGVACKRLGTIIDDAAGEAFDKIAKS 183

OY 61 LSLIHPECSMTSGGKAITHLAKQGRFHFDIKPELHAKKCDSPFGQLQVHTDKIMKK 120

DB 184 LGL-PYP-----GGALKEFLAVGSDPTRYALPRALIGRKDCDSFSLKTPAAARIATLT 236

OY 121 EKESEIEKQILSSAADIAATVQHTMACHLVKTHTRAIILFKQKDLIFQNNAVLVAAGGV 180

DB 237 TTDD-----ARRDLAAGVQAAIARQLSERVDRAMKLYK--DSHPEDLRFVYAGGV 285

OY 181 ASNFYIRALIELITNATQCTLLCPPRCLCTDNGIMAMNGIERLRAGIGILHDIIGIRYE 240

DB 286 AANGAVRALLLADCEKNGFSFAPPLAYCTDNAMAMIALAGERL--ALGIFDDIDATA-R 342

OY 241 PKCPL 245

DB 343 PRWPL 347

RESULT 8

H83572

O-sialoglycoprotein endopeptidase PA0580 [imported] - Pseudomonas aeruginosa (strain PAO

C:Species: Pseudomonas aeruginosa

C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C:Accession: H83572

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Micozuchi, S.D.; Warrenner, P.; Hickey, M.J.; Bradman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lattig, K.; Lim, J.; Lory, S.; Olson, M.V. Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pathob

A:Reference number: A82950; MUID:20437337; PMID:10984043

A:Accession: H83572

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-341 <STO>

A:Cross-references: GB:AE004494; GB:AE004091; NID:G9946446; PIDN:AG03969.1; GSPDB:GN001

A:Experimental source: strain PAO1

C:Genetics:

A:Gene: gcp

C:Superfamily: O-sialoglycoprotein endopeptidase

Query Match 21.0%; Score 291.5; DB 2; Length 341;

Best Local Similarity 35.4%; Pred. No. 1.5e-18;

Matches 87; Conservative 37; Mismatches 99; Indels 23; Gaps 9;

QY 1 MEAHALTIRLTNK-VEPPFVLVLLISGGHCLALVQGVSDPFLILGKSLDIAFGMDLKVAR 59
 Db 112 MEGHLAPMLEEQPRPFVALLVSGHTQIVRVGIGRYQLLGSSVDDAAGEAPDKTX 171
 QY 60 RLSLIKPECSMTSGGKALIEHLAKQGNRPHFDIKPPLHAKNCDPSFTGLQHTVDKTIIMK 119
 Db 172 LIGL-GYP-----GGPEIATLARSGTPGRFVFPMPDRPQDLFSPGLKTFITLN-TWQ 223
 QY 120 KEKEEGIEKGQILSSAADIAATVOHTMACHLVKTRTRAILFCQKRDLLPQNNAVIVASGG 179
 Db 224 RCEVAGDSEQ---TRCDIALAFQAVVETLTKCRRL---KQYGL--KN--LVVAGG 272
 QY 180 VASNFYIRALIEILTNAQCTLLCPPRILCTDNGIMIMANGIERLAGIGLHDIEGIRY 239
 Db 273 VSNQALRSGLERKLGEMKGVFARPRFCTDNGMIMIVAGCQRLLAG---QHDGPAISV 329
 QY 240 EPKCP 245
 Db 330 QPRWPM 335

RESULT 9

sialoglycoproteinase - Aquifex aeolicus
 C:Species: Aquifex aeolicus
 C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 09-Jul-2004
 C:Accession: G70369
 R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O.V.
 Nature 392, 353-358, 1998
 A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
 A:Reference number: A70300; MUID:98196666; PMID:9537320
 A:Accession: G70369
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-335 <AOE>
 A:Cross-references: UNIPROT:O66986; GB:AB000708; NID:g2983356; PIDN:AAC06951.1; PID:g298
 A:Experimental source: strain VFS
 C:Genetics:
 A:Gene: gcp
 C:Superfamily: O-sialoglycoprotein endopeptidase

Query Match 21.0%; Score 290.5; DB 2; Length 335;
 Best Local Similarity 34.3%; Pred. No. 1.8e-18;
 Matches 85; Conservative 46; Mismatches 82; Indels 35; Gaps 10;
 QY 1 MEAHALTIRLTNKVEPPFVLVLLISGGHCLALVQGVSDPFLILGKSLDIAFGMDLKVAR 60
 Db 112 MEGHLAPMLEEQPRPFVALLVSGHTQIVRVGIGRYQLLGSSVDDAAGEAPDKTX 171
 QY 61 RLSLIKPECSMTSGGKALIEHLAKQGNRPHFDIKPPLHAKNCDPSFTGLQHTVDKTIIMK 120
 Db 172 LIGL-GYP-----GGPEIATLARSGTPGRFVFPMPDRPQDLFSPGLKTFITLN-TWQ 223
 QY 121 KEKEEGIEKGQILSSAADIAATVOHTMACHLVKTRTRAILFCQKRDLLPQNNAVIVASGG 180
 Db 224 RCEVAGDSEQ---TRCDIALAFQAVVETLTKCRRL---KQYGL--KN--LVVAGG 272
 QY 181 VASNFYIRALIEILTNAQCTLLCPPRILCTDNGIMIMANGIERLAGIGLHDIEGIRY 239
 Db 266 VASNFYIRALIEILTNAQCTLLCPPRILCTDNGIMIMANGIERLAGIGLHDIEGIRY 239
 QY 238 RYBCKP 245
 Db 322 --QPNIFL 327

RESULT 10

H64074
 O-sialoglycoprotein endopeptidase (BC 3.4.24.57) - Haemophilus influenzae (strain Rd KW2
 N:Alternate names: sialoglycoproteinase
 C:Species: Haemophilus influenzae

C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 09-Jul-2004
 C:Accession: H64074
 R:Feischmann, A.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kierlavage, A.
 ; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.
 , D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhmann, J.L.; Geoghegan, N.S.M.
 Science 269, 496-512, 1995
 A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,
 A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
 A:Reference number: A64000; MUID:95350630; PMID:7542800
 A:Accession: H64074
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-342 <TIGR>
 A:Cross-references: UNIPROT:P43764; GB:U32735; GB:I42023; NID:g1573509; PIDN:AAC22187.1;
 C:Superfamily: O-sialoglycoprotein endopeptidase
 C:Keywords: hydrolase; metalloproteinase

Query Match 20.9%; Score 289.5; DB 2; Length 342;
 Best Local Similarity 35.7%; Pred. No. 2.3e-18;
 Matches 82; Conservative 30; Mismatches 95; Indels 23; Gaps 7;

QY 1 MEAHALTIRLTNKVEPPFVLVLLISGGHCLALVQGVSDPFLILGKSLDIAFGMDLKVAR 59
 Db 112 MEGHLAPMLEEQPRPFVALLVSGHTQIVRVGIGRYQLLGSSVDDAAGEAPDKTX 171
 QY 60 RLSLIKPECSMTSGGKALIEHLAKQGNRPHFDIKPPLHAKNCDPSFTGLQHTVDKTIIMK 119
 Db 172 LIGL-DYP-----GGAALSRLEKCTPRPFPRPMTBAGIDFSGKTFPAANTVNG 224
 QY 120 KEKEEG--IEKGQILSSAADIAATVOHTMACHLVKTRTRAILFCQKRDLLPQNNAVIVAS 177
 Db 225 AIKNGELIEQ-----TKADIAVAFQDAVVDILA-----IKCK-RALKETGYKRLVIA 271
 QY 178 GGVASNFYIRALIEILTNAQCTLLCPPRILCTDNGIMIMANGIERLAG 227
 Db 272 GGVASNFYIRALIEILTNAQCTLLCPPRILCTDNGIMIMANGIERLAG 227

RESULT 11

A10079
 Probable glycoproteinase gcp [imported] - Yersinia pestis (strain CO92)
 C:Species: Yersinia pestis
 C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Nov-2001
 C:Accession: A10079
 R:Parhill, J.; Wren, B.W.; Thomson, N.R.; Tibball, R.W.; Holden, M.T.G.; Prentice, M.B.
 ; deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
 11, M.; Rutherford, K.; Simmonds, W.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrett,
 Nature 413, 523-527, 2001
 A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
 A:Reference number: AB0001; MUID:21470413; PMID:11586360
 A:Accession: A10079
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-337 <KUR>
 A:Cross-references: GB:AL590842; PIDN:CAC89500.1; PID:g15978736; GSPDB:GN00175
 C:Genetics:
 A:Gene: gcp
 C:Superfamily: O-sialoglycoprotein endopeptidase

Query Match 20.2%; Score 279.5; DB 2; Length 337;
 Best Local Similarity 32.8%; Pred. No. 1.8e-17;
 Matches 81; Conservative 35; Mismatches 102; Indels 29; Gaps 7;
 QY 1 MEAHALTIRLTNKVEPPFVLVLLISGGHCLALVQGVSDPFLILGKSLDIAFGMDLKVAR 59
 Db 112 MEGHLAPMLEEQPRPFVALLVSGHTQIVRVGIGRYQLLGSSVDDAAGEAPDKTX 171
 QY 60 RLSLIKPECSMTSGGKALIEHLAKQGNRPHFDIKPPLHAKNCDPSFTGLQHTVDKTIIMK 119
 Db 172 LIGL-DYP-----GGPEIATLARSGTPGRFVFPMPDRPQDLFSPGLKTFITLN-TWQ 223
 QY 120 KEKEEGIEKGQILSSAADIAATVOHTMACHLVKTRTRAILFCQKRDLLPQNNAVIVASG 178
 Db 224 RCEVAGDSEQ---TRCDIALAFQAVVETLTKCRRL---KQYGL--KN--LVVAGG 272

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2005, 15:16:01 ; Search time 64.7185 Seconds
(without alignments)
2112.614 Million cell updates/sec

Title: US-10-649-273-2_COPY_148_414
Perfect score: 1385
Sequence: 1 MESHALLTRLNKVEPFLV.....DISKEVGASIKVQLKMEI 267

Scoring table: BL0SUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt_03:*
1: uniProt_sprot:*
2: uniProt_trembl:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1385	100.0	414	2	096EV9
2	1358.5	98.1	439	2	09H4B0
3	1231	88.9	414	2	06PEB4
4	1224	88.4	414	2	08BLB6
5	1216	87.8	414	2	09DON0
6	1116	80.6	467	2	06AYN7
7	1105	79.8	364	2	096NH5
8	859	62.0	404	2	08JFW3
9	855	61.7	404	2	08JFR7
10	453.5	32.7	401	2	07O9I8
11	415.5	30.0	323	2	09W0S6
12	415.5	30.0	409	2	09WMD6
13	407.5	29.4	480	2	022145
14	369	26.6	335	2	073H71
15	356.5	25.7	360	2	092I18
16	349	25.4	362	2	098E16
17	342	25.2	387	2	068XR3
18	342	24.7	251	2	093FU2
19	339.5	24.5	389	2	06ND54
20	338	24.4	365	2	08UC47
21	338	24.4	366	2	07CMW8
22	338	24.4	387	1	GCP_RICPR
23	334.5	24.2	359	2	08YUJ1
24	334.5	24.2	359	2	08FYI5
25	333.5	24.1	344	2	07PAG7
26	329.5	23.8	344	2	092UX6
27	326.5	23.6	357	2	089W11
28	326.5	23.6	357	2	G5A25
29	321.5	23.2	346	2	07VXN4
30	320	23.1	364	2	06G1R3
31	317	22.9	340	2	06FCR9

32	316.5	22.9	346	2	07M668	07M668 bordetella
33	315.5	22.8	346	2	07W134	07W134 bordetella
34	304.5	22.0	343	2	09CLJ1	09CLJ1 pasteurella
35	304	21.9	353	2	07VQ09	07VQ09 candidatus
36	302.5	21.8	341	2	07NUE3	07NUE3 chromobacter
37	302	21.8	339	2	06LV10	06LV10 photobacter
38	299.5	21.6	255	2	0677H2	0677H2 hyacinthus
39	298	21.5	364	2	06FYF1	06FYF1 bartonella
40	293	21.2	344	2	06SRP0	06SRP0 pseudomonas
41	291.5	21.0	341	2	09ISV7	09ISV7 pseudomonas
42	290.5	21.0	335	1	GCP_AOUAE	06986 aquifex aeo
43	289.5	20.9	342	1	GCP_HAEIN	043764 haemophilus
44	286.5	20.7	337	2	082XN2	082XN2 nitrosomona
45	285.5	20.6	348	2	09L7A5	09L7A5 haemophilus

ALIGNMENTS

```
RESULT 1
096EV9 PRELIMINARY: PRT; 414 AA.
AC 096EV9;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE OSGBPL1 protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Burow K.H., Scheffer C.F., Hsieh F.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loguettano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.N., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Scherch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RA Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC011904; AAH11904.1; -
DR MEROPS; M22.004; -
DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro: IPR00905; Peptidase M22.
DR InterPro: IPR009180; Pept M22_Osialgl.
DR Pfam: PF00814; Peptidase M22_1.
DR PIRSF: PIRSF004537; Osialglc_Optds; 1.
DR PRINTS: PR00789; Osialglc_Optds; 1.
DR ProDom: PD002367; Peptidase M22; 1.
DR TIGFAMS: TIGR00329; GCP; 1.
DR SEQUENCE 414 AA; 45122 MW; A536B333F5C6B8DD CRC64;
```

Query Match 100.0%; Score 1385; DB 2; Length 414;
Best Local Similarity 100.0%; Pred. No. 2,9e-110;
Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEAHALTRLTNKKVEPFLVLLISGGHCLLALVQVSPDFLLGKSLDIAFGMDLKVARR 60
DB 148 MEAHALTRLTNKKVEPFLVLLISGGHCLLALVQVSPDFLLGKSLDIAFGMDLKVARR 207
QY 61 LSLIKHEPCSTMSGKAIIEHLAKQGNRPHFDIKPPLHAKNCDSFPTGLQHTDXTKIMKK 120
DB 208 LSLIKHEPCSTMSGKAIIEHLAKQGNRPHFDIKPPLHAKNCDSFPTGLQHTDXTKIMKK 267

QY 121 EKEGIEKGQILSSAADIAATVQHTMACHLVKTTRAILFCQKRDLPQNNNAVLAASGV 180
DB 268 EKEGIEKGQILSSAADIAATVQHTMACHLVKTTRAILFCQKRDLPQNNNAVLAASGV 327

QY 181 ASNFYIRRALIEILTNATQCTLLCPPRLCTDNGMIAMNGIERLRAGLGIHDIGIRYE 240
DB 328 ASNFYIRRALIEILTNATQCTLLCPPRLCTDNGMIAMNGIERLRAGLGIHDIGIRYE 387

QY 241 PKCPLGVDISKVEGASIKVPLKMEI 267
DB 388 PKCPLGVDISKVEGASIKVPLKMEI 414

RESULT 2

Q9H4B0 PRELIMINARY; PRT; 439 AA.
ID Q9H4B0
AC Q9H4B0; 2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DE Putative sialoglycoprotein type 2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Uterus;
RA Chen J.M., Fortunato M., Barrett A.U.;
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ295148; CAC14666.1; -
DR MEROPS: M22.004; -
DR Genew: HGNC:23075; OSGEPL1.
DR GO: GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.
DR GO: GO:0008233; F:peptidase activity; IEA.
DR GO: GO:0008270; F:zinc ion binding; IEA.
DR GO: GO:000508; P:proteolysis and peptidolysis; IEA.
DR InterPro: IPR000905; Peptidase_M22.
DR InterPro: IPR009180; Pept_M22_Osialgl.
DR Pfam: PF00814; Peptidase_M22; 1.
DR PIRSF: PIRSF004537; Osialglc_ptids; 1.
DR PRINTS: PR00789; OSIALOPTASE.
DR Prodom: PD002367; Peptidase_M22; 1.
DR TIGRFAMs: TIGR00329; gcp; 1.
KW Protease.
SQ SEQUENCE 439 AA; 48040 MW; 44849372C784E41F CRC64;

Query Match 98.1%; Score 1358.5; DB 2; Length 439;
Best Local Similarity 91.1%; Pred. No. 5.8e-106;
Matches 266; Conservative 0; Mismatches 1; Indels 25; Gaps 1;

QY 1 MEAHALTRLTNKKVEPFLVLLISGGHCLLALVQVSPDFLLGKSLDIAFGMDLKVARR 60
DB 148 MEAHALTRLTNKKVEPFLVLLISGGHCLLALVQVSPDFLLGKSLDIAFGMDLKVARR 207
QY 61 LSLIKHEPCSTMSGKAIIEHLAKQGNRPHFDIKPPLHAKNCDSFPTGLQHTDXTKIMKK 120
DB 208 LSLIKHEPCSTMSGKAIIEHLAKQGNRPHFDIKPPLHAKNCDSFPTGLQHTDXTKIMKK 267
QY 121 EKEGI-----EKGQILSSAADIAATVQHTMACHLVKT 155

DB 268 EKEGIEKIFLSKVEQINIPGLCIKIAHFCRYEKQILSSAADIAATVQHTMACHLVKT 327
QY 156 RALIFCKQRDLPPQNNNAVLAASGVASNFYIRRALIEILTNATQCTLLCPPRLCTDNGMI 215
DB 328 RALIFCKQRDLPPQNNNAVLAASGVASNFYIRRALIEILTNATQCTLLCPPRLCTDNGMI 387

QY 216 IAMNGIERLRAGLGIHDIGIRYEPKCPGLGVDISKVEGASIKVPLKMEI 267
DB 388 IAMNGIERLRAGLGIHDIGIRYEPKCPGLGVDISKVEGASIKVPLKMEI 439

RESULT 3

O6PEB4 PRELIMINARY; PRT; 414 AA.
ID O6PEB4
AC O6PEB4;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CZECH II; TISSUE=Mammary tumor;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Scheffer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Millar S.J.,
RA Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywicki M.J., Skalska U., Smalins D.E., Scherch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CZECH II; TISSUE=Mammary tumor;
RA Strausberg R.;
RL Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC058172; AAH58172.1; -
DR GO: GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.
DR GO: GO:0008270; F:zinc ion binding; IEA.
DR GO: GO:000508; P:proteolysis and peptidolysis; IEA.
DR InterPro: IPR000905; Peptidase_M22.
DR InterPro: IPR009180; Pept_M22_Osialgl.
DR Pfam: PF00814; Peptidase_M22; 1.
DR PIRSF: PIRSF004537; Osialglc_ptids; 1.
DR PRINTS: PR00789; OSIALOPTASE.
DR Prodom: PD002367; Peptidase_M22; 1.
DR TIGRFAMs: TIGR00329; gcp; 1.
KW Hypothetical protein.
SQ SEQUENCE 414 AA; 44962 MW; 0418AD93C6FEBCAE CRC64;

Query Match 88.9%; Score 1231; DB 2; Length 414;
Best Local Similarity 87.3%; Pred. No. 4.6e-97;
Matches 233; Conservative 15; Mismatches 19; Indels 0; Gaps 0;

QY 1 MEAHALTRLTNKKVEPFLVLLISGGHCLLALVQVSPDFLLGKSLDIAFGMDLKVARR 60

Db 148 MEAHALTRLRNKNKFFPLVLLISGHCILALVGVSPFLILGKSLDIAPGMDLKVARR 207
Qy 61 LSLIKHEPCSTMSGGKAIIEHLAKQGNRPHFDIKPPLHAKNCDFSTGLQHVTDKIIMKK 120
Db 208 LSLIKHEPCSTMSGGKAIIEHLAKQGNRPHFTINPMQAKNCDFSTGLQHVTDKIIMKK 267
Qy 121 EKEBIEKGQILSSAADIAAVQHTACHLAKRTHRALTFCKQKVLSPANAVLVASGV 180
Db 268 EKEBIEKGQILSSAADIAAVQHTACHLAKRTHRALTFCKQKVLSPANAVLVASGV 327
Qy 181 ASNFYIRALREILTNATQCTLLCPPRLCTDNGMIANNGIERLRAGLIDIEGIRYE 240
Db 328 ASNLTYIRKALREIVANATQCTLLCPPRLCTDNGMIANNGIERLRAGLIDIEGIRYE 387
Qy 241 PKCPLGVDISKEVGEASIKVPOLKMEI 267
Db 388 PKCPLGVDISREVAALIKVRLKVAL 414
RESULT 4
Q8BLB6 PRELIMINARY; PRT; 414 AA.
AC 08BLB6;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Mus musculus adult male corpora quadrigemina cDNA, RIKEN full-length
enriched library, clone:B30219017 product:similar to PUTATIVE
SIALOGLYCOPROTEASE TYPE 2.
GN Name=Osgp11;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;
RX MEDLINE=9279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;
RA The FANTOM Consortium,
the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa S., Nagasaki S., Sasaki N., Carninci P.,
Kono H., Akiyama J., Nishi K., Kitenuma T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
Yamamoto R., Matsunoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,

RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;
RA Adachi N., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hirooka T., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Katoh H., Kawai J., Kojima Y., Kondo S., Kono H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA Saito R., Satoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK045669; BAC32450.1; -.
DR MEROPS: M22.004; -.
DR MGI: MGI:1919335; Osgp11.
DR GO: GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.
DR GO: GO:0008233; F:peptidase activity; IEA.
DR GO: GO:0008270; F:zinc ion binding; IEA.
DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro: IPR000905; Peptidase_M22.
DR InterPro: IPR009180; Pept_M22_Osialgl.
DR Pfam: PF00814; Peptidase_M22; 1.
DR PIRSF: PIRSF004537; Osialglc_Pptds; 1.
DR PRINTS: PRO0789; OSIALOPTASE_
DR Prodom: PD002367; Peptidase_M22; 1.
DR TIGRFAMs: TIGR00329; gcp; 1.
KM Protease.
SQ SEQUENCE 414 AA; 44927 MW; 9A2D27B576C7DB24 CRC64;
Query Match 88.4%; Score 1224; DB 2; Length 414;
Best Local Similarity 87.3%; Pred. No. 1.8e-96;
Matches 233; Conservative 14; Mismatches 20; Indels 0; Gaps 0;
Qy 1 MEAHALTRLRNKNKFFPLVLLISGHCILALVGVSPFLILGKSLDIAPGMDLKVARR 60
Db 148 MEAHALTRLRNKNKFFPLVLLISGHCILALVGVSPFLILGKSLDIAPGMDLKVARR 207
Qy 61 LSLIKHEPCSTMSGGKAIIEHLAKQGNRPHFDIKPPLHAKNCDFSTGLQHVTDKIIMKK 120
Db 208 LSLIKHEPCSTMSGGKAIIEHLAKQGNRPHFTINPMQAKNCDFSTGLQHVTDKIIMKK 267
Qy 121 EKEBIEKGQILSSAADIAAVQHTACHLAKRTHRALTFCKQKVLSPANAVLVASGV 180
Db 268 EKEBIEKGQILSSAADIAAVQHTACHLAKRTHRALTFCKQKVLSPANAVLVASGV 327
Qy 181 ASNFYIRALREILTNATQCTLLCPPRLCTDNGMIANNGIERLRAGLIDIEGIRYE 240
Db 328 ASNLTYIRKALREIVANATQCTLLCPPRLCTDNGMIANNGIERLRAGLIDIEGIRYE 387
Qy 241 PKCPLGVDISKEVGEASIKVPOLKMEI 267
Db 388 PKCPLGVDISREVAALIKVRLKVAL 414
RESULT 5
Q9DON0 PRELIMINARY; PRT; 414 AA.
AC Q9DON0;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Mus musculus 10 days embryo whole body cDNA, RIKEN full-length
enriched library, clone:2610001M19 product:similar to PUTATIVE
SIALOGLYCOPROTEASE TYPE 2.

GN Name=Osgp11; (Mouse).
OS Mus musculus.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=whole body;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=whole body;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=whole body;
RA The FANTOM Consortium,
RT the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=whole body;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=whole body;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Kono H., Akiyama J., Nishi K., Kitsuwa T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsunoto H., Sakaguchi S., Ikegami T., Kasahagi K.,
RA Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohara E., Wachihi M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai U.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=whole body;
RA Adachi T., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
RA Arawaka T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
RA Hanagawa T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,
RA Imotani K., Ishii Y., Itoh M., Izawa M., Kakuwa T., Kato H.,
RA Kawai U., Kojima Y., Kono H., Kouda M., Koya S., Kurihara C.,
RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
RA Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,
RA Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,
RA Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
RA Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
RA Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AK011265; BAB27506.1; -
DR MEROPS; M22.004; -
DR MGD; MGI:1919335; Osgp11.
DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0006508; P:proteolysis; P:peptidolysis; IEA.
DR InterPro; IPR000905; Peptidase_M22.

DR InterPro; IPR009180; Pept_M22_Osialgl.
DR Pfam; PF00814; Peptidase_M22; 1.
DR PIRSF; PIRSF04537; Osialglc_ptide; 1.
DR PRINTS; PR00789; OSIALOPTASE.
DR ProDom; PD002367; Peptidase_M22; 1.
DR TIGRFAMs; TIGR00329; gcp; 1.
KM Peptidase.
SQ SEQUENCE 414 AA; 44999 MW; 999BC689944DB24 CRC64;
Query Match 87.8%; Score 1216; DB 2; Length 414;
Best Local Similarity 86.9%; Pred. No. 8.9e-96;
Matches 232; Conservative 14; Mismatches 21; Indels 0; Gaps 0;
QY 1 MEAHLITRLTNKVEFPPLVLLISGGHLLAVGVSDPLLKSLDIAPGMDLVAR 60
DB 148 MEAHLITRLTNKVEFPPLVLLISGGHLLAVGVSDPLLKSLDIAPGMDLVAR 207
QY 61 LSLIHPECSITMSGKALIEHLAKGNRRFPDIPKPLHNAKNCDFSTGQHTDKIMK 120
DB 208 LSLIHPECSITMSGKALIEHLAKGNRRFPDIPKPLHNAKNCDFSTGQHTDKIMK 267
QY 121 EKEEGIEKGQILSSADIAATVQHTMACTLVKTRTRAILFCORDLIPONNAVVASGV 180
DB 268 EKEEGIEKGQILSSADIAAAYQHTACTHARTTRAILFCOKILSPANAVLVSGV 327
QY 181 ASNPYIRRALIELITNATQCTLLCPPRLCTDNGIMIANWGIRLRAGLIDIGIRYE 240
DB 328 ASNPYIRRALIELITNATQCTLLCPPRLCTDNGIMIANWGIRLRAGLIDIGIRYE 387
QY 241 PKCPGVDSISKEVGASIKVPOLKMEI 267
DB 388 PKCPGVDSISKEVGASIKVPOLKMEI 414
RESULT 6
Q6AYN7 PRELIMINARY; PRT; 467 AA.
AC Q6AYN7;
DT 25-OCT-2004 (TREMBLrel. 28, Created)
DT 25-OCT-2004 (TREMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TREMBLrel. 28, Last annotation update)
DE Hypothetical protein (Fragment).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schler G.D.,
RA Altshul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diachenko L., Marusina K., Farmer A.C., Rubin G.M., Hong L.,
RA Stalcenon M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loggellano N.A., Peters G.J., Abramson R.D., Mollady S.J.,
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A.C., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green B.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalls D.E., Scherch A., Schein J.E.,
RT Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;

RA Director MGC Project;
 RU Submitted (Aug-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC078974; AAH78974.1;
 DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.
 DR GO; GO:0008270; F:zinc ion binding; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000905; Peptidase_M22.
 DR InterPro; IPR001807; Pept_M22_Osialgl.
 DR InterPro; IPR002016; Peroxidase.
 DR Pfam; PF00814; Peptidase_M22; 1.
 DR PIRSF; PIRSF004537; Osialgl_ptds; 1.
 DR PRINTS; PR00789; OSIALOPTASE.
 DR Prodom; PD002367; Peptidase_M22; 1.
 DR TIGRFAMs; TIGR00329; gcp; 1.
 DR PROSITE; PS00435; PEROXIDASE_1; UNKNOWN_1.
 KW Hypothetical protein.
 FT NON TER 467
 SQ SEQUENCE 467 AA; 50799 MW; 474B1B1959B8AC0 CRC64;
 Query Match 80.6%; Score 1116; DB 2; Length 467;
 Best Local Similarity 88.4%; Pred. No. 3.8e-87;
 Matches 214; Conservative 10; Mismatches 18; Indels 0; Gaps 0;
 QY 1 MEAHLITRLTNKVEPFLVLLISGHCCLALVGVSPFLLLGKSLDIAPGMDLVARR 60
 DB 148 MEAHLITRLTNKVEPFLVLLISGHCCLALVGVSPFLLLGKSLDIAPGMDLVARR 207
 QY 61 LSLIKPECSMGSGKAIIEHLAKQGNRFFDIKPLHNAKNCDFSTGLQHTVDKTIIMK 120
 DB 208 LSLIKPECSMGSGKAIIEHLAKQGNRFFDIKPLHNAKNCDFSTGLQHTVDKTIIMK 267
 QY 121 EKEGIEKGQILSSADIAATVQHTMACLVKTRTHRALIFCKORDLPONNAVLVASGV 180
 DB 268 EKEGIEKGQILSSADIAATVQHTMACLVKTRTHRALIFCKORDLPONNAVLVASGV 327
 QY 181 ASNFYIRALBELTNATQCTLLCPPRLCTDNGIMIANNGIERLACGILHDIGIRYE 240
 DB 328 ASNFYIRALBELTNATQCTLLCPPRLCTDNGIMIANNGIERLACGILHDIGIRYE 387
 QY 241 PK 242
 DB 388 PK 389
 RESULT 7
 Q96NH5 PRELIMINARY; PRT; 364 AA.
 ID Q96NH5
 AC Q96NH5;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Hypothetical protein FLJ30879.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX PubMed=14702039; DOI=10.1038/ng1285;
 RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
 RA Makatsuta A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
 RA Sekine M., Ohtsushi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
 RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahara K.,
 RA Murakami K., Yaizu T., Iwayanagi T., Wagatsuma M., Shiratori A.,
 RA Takahashi K., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,
 RA Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,
 RA Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,
 RA Tanai H., Kimura M., Watanabe M., Hirotsuka S., Chiba Y., Ishida S.,
 RA Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J.,
 RA Kanehori K., Takahashi-Fuji A., Hara R., Tanase T., Nomura Y.,
 RA Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,

RA Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,
 RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,
 RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
 RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
 RA Hisigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
 RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
 RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
 RA Ono T., Yamada K., Fujii Y., Ozaki K., Hiroo M., Ohmori Y.,
 RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
 RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
 RA Matsumura K., Nakajima Y., Mizuno T., Motomaga M., Sasaki M.,
 RA Togsushi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
 RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
 RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuno Y., Yamashita R.,
 RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.,
 RT "Complete sequencing and characterization of 21,243 full-length human
 RT cDNAs".
 RU Nat. Genet. 36:40-45(2004).
 DR EMBL; AK055441; BAB70923.1; -.
 DR MEROPS; M22.004; -.
 DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.
 DR GO; GO:0008270; F:peptidase activity; IEA.
 DR GO; GO:0008270; F:zinc ion binding; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000905; Peptidase_M22.
 DR InterPro; IPR009180; Pept_M22_Osialgl.
 DR Pfam; PF00814; Peptidase_M22; 1.
 DR PIRSF; PIRSF004537; Osialgl_ptds; 1.
 DR PRINTS; PR00789; OSIALOPTASE.
 DR Prodom; PD002367; Peptidase_M22; 1.
 DR TIGRFAMs; TIGR00329; gcp; 1.
 KW Protease.
 SQ SEQUENCE 364 AA; 39528 MW; E0B605A07D0EC3D6 CRC64;
 Query Match 79.8%; Score 1105; DB 2; Length 364;
 Best Local Similarity 98.6%; Pred. No. 2.5e-86;
 Matches 214; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 MEAHLITRLTNKVEPFLVLLISGHCCLALVGVSPFLLLGKSLDIAPGMDLVARR 60
 DB 148 MEAHLITRLTNKVEPFLVLLISGHCCLALVGVSPFLLLGKSLDIAPGMDLVARR 207
 QY 61 LSLIKPECSMGSGKAIIEHLAKQGNRFFDIKPLHNAKNCDFSTGLQHTVDKTIIMK 120
 DB 208 LSLIKPECSMGSGKAIIEHLAKQGNRFFDIKPLHNAKNCDFSTGLQHTVDKTIIMK 267
 QY 121 EKEGIEKGQILSSADIAATVQHTMACLVKTRTHRALIFCKORDLPONNAVLVASGV 180
 DB 268 EKEGIEKGQILSSADIAATVQHTMACLVKTRTHRALIFCKORDLPONNAVLVASGV 327
 QY 181 ASNFYIRALBELTNATQCTLLCPPRLCTDNGIMIA 217
 DB 328 ASNFYIRALBELTNATQCTLLCPPRLCTDNGIMIA 364
 RESULT 8
 Q8UFW3 PRELIMINARY; PRT; 404 AA.
 ID Q8UFW3
 AC Q8UFW3;
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE S1:dz211013.4 (Novel glycoprotease).
 GN Name=dz211013.4;
 OS Brachydanio rerio (Zebrafish) (Danio rerio).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Danio.
 NCBI_TaxID=7955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC Babbage A.;
 RA Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.

DR EMBL; AL591593; CAD43471.1; -.
DR MEROPS; M22.004; -.
DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0006508; F:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009180; Peptidase M22.
DR InterPro; IPR009180; Pept M22_Osialgl.
DR Pfam; PF00814; Peptidase M22; 1.
DR PIRSF; PIRSF004537; Osialglc_ptide; 1.
DR PRINTS; PR00789; OSIALOPTASE.
DR ProDom; PD002367; Peptidase_M22; 1.
DR TIGRFAMs; TIGR00329; gcp; 1.
KM Protease.
SQ SEQUENCE 404 AA; 43956 MW; 3A6B1DD050737E35 CRC64;
Query Match 62.0%; Score 859; DB 2; Length 404;
Best Local Similarity 62.3%; Pred. No. 3,4e-65;
Matches 165; Conservative 38; Mismatches 62; Indels 0; Gaps 0;
QY 1 MEAHLTRITNKVEPFLVLLISGGHCLALVGVSDFLIGKSLDIAPGMDLVKVR 60
DB 136 MEAHLTVRMPLPDPFPLVLLVSGHSLALAKGIDFLLIGQLDEAAGDTLXKARR 195
QY 61 LSLIKHECSTWSGKATIEHLAKOGNRFHPDIKPLHAKNCDSPFTGLQHTVDKIIMKK 120
DB 196 LSLJNHPECSTLSSGQALIERLAKGSDRLAFHISPWGQNYDCNFPAGLRTOITGAINKK 255
QY 121 EKEGIEKGQILSSADIAATVQHTMACHLVKTRTRAILFCCKORDLPONNAVLVAGCV 180
DB 256 EKEGVEVGQPLSCVKDIAAASQHTVASHLAKRTTRAILFCCKSLDEQNPFTLIVSGGV 315
QY 181 ASNFYIRALIELTNATQCTLLCPPELCTDNGMIAMNGIERLRAGLIDIGIRYE 240
DB 316 ASNEYIRQLKITDATTGLHLCPSPKCTDNGMIAMNGIERLRKQKGLISYSEVSVE 375
QY 241 PKCPLGVDSIKVEGASIKVPOLKM 265
DB 376 PKAPLGIDITSEVKEAIVKPKKL 400

RESULT 9
Q8JFR7 PRELIMINARY; PRT; 404 AA.
AC Q8JFR7;
DT 01-OCT-2002 (Tremblrel. 22, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE SI:d72B14.6 (Novel glycoprotease).
GN Name=SI:d72B14.6;
OS Brachydanio rerio (zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Clark G.;
RL Submitted (JUN-2003) to the EMBL/Genbank/DBJ databases.
DR EMBL; AL672217; CAD43443.1; -.
DR MEROPS; M22.004; -.
DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0006508; F:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009180; Peptidase M22.
DR InterPro; IPR009180; Pept M22_Osialgl.
DR Pfam; PF00814; Peptidase_M22; 1.
DR PIRSF; PIRSF004537; Osialglc_ptide; 1.
DR PRINTS; PR00789; OSIALOPTASE.
DR ProDom; PD002367; Peptidase_M22; 1.
DR TIGRFAMs; TIGR00329; gcp; 1.
KM Protease.

SQ SEQUENCE 404 AA; 44027 MW; 6FE98653A651860F CRC64;
Query Match 61.7%; Score 855; DB 2; Length 404;
Best Local Similarity 61.9%; Pred. No. 7,6e-65;
Matches 164; Conservative 39; Mismatches 62; Indels 0; Gaps 0;
QY 1 MEAHLTRITNKVEPFLVLLISGGHCLALVGVSDFLIGKSLDIAPGMDLVKVR 60
DB 136 MEAHLTVRMPLPDPFPLVLLVSGHSLALAKGIDFLLIGQLDEAAGDTLXKARR 195
QY 61 LSLIKHECSTWSGKATIEHLAKOGNRFHPDIKPLHAKNCDSPFTGLQHTVDKIIMKK 120
DB 196 LSLJNHPECSTLSSGQALIERLAKGSDRLAFHISPWGQNYDCNFPAGLRTOITGAINKK 255
QY 121 EKEGIEKGQILSSADIAATVQHTMACHLVKTRTRAILFCCKORDLPONNAVLVAGCV 180
DB 256 EKEGVEVGQPLSCVKDIAAASQHTVASHLAKRTTRAILFCCKSLDEQNPFTLIVSGGV 315
QY 181 ASNFYIRALIELTNATQCTLLCPPELCTDNGMIAMNGIERLRAGLIDIGIRYE 240
DB 316 ASNEYIRQLKITDATTGLHLCPSPKCTDNGMIAMNGIERLRKQKGLISYSEVSVE 375
QY 241 PKCPLGVDSIKVEGASIKVPOLKM 265
DB 376 PKAPLGIDITSEVKEAIVKPKKL 400

RESULT 10
Q7Q918 PRELIMINARY; PRT; 401 AA.
ID Q7Q918;
AC Q7Q918;
DT 01-MAR-2004 (Tremblrel. 26, Created)
DT 01-MAR-2004 (Tremblrel. 26, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE AgCP14990 (Fragment).
GN Name=agcG46164; ORFNames=ENSANG0000007922;
OS Anopheles gambiae str. BEST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.
OX NCBI_TaxID=180454;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=BEST;
RC Anopheles Genome Sequencing Consortium;
RL Submitted (MAR-2002) to the EMBL/Genbank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/Genbank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAB01008900; BAA09387.1; -.
DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0006508; F:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009180; Peptidase_M22.
DR InterPro; IPR009180; Pept M22_Osialgl.
DR Pfam; PF00814; Peptidase M22; 1.
DR PIRSF; PIRSF004537; Osialglc_ptide; 1.
DR PRINTS; PR00789; OSIALOPTASE.
DR ProDom; PD002367; Peptidase_M22; 1.
DR TIGRFAMs; TIGR00329; gcp; 1.
FT NON TER 1
SQ SEQUENCE 401 AA; 43849 MW; CC9426723D3FD4F1 CRC64;
Query Match 32.7%; Score 453.5; DB 2; Length 401;
Best Local Similarity 38.4%; Pred. No. 1,9e-30;
Matches 106; Conservative 50; Mismatches 109; Indels 11; Gaps 4;
QY 1 MEAHLTRITNKVEPFLVLLISGGHCLALVGVSDFLIGKSLDIAPGMDLVKVR 60
DB 119 MOAHLAMRWITITIPFLCLVSGHSLVIVESTARRLGLGTLDDPAGMLDKIARR 178
QY 61 LSLIKHECSTWSGKATIEHLAKOG-----NRFHPDIKPLHAKNCDSPFTGLQHTVDK 115
DB 179 LKLRVAVAKYAGMSGGATLAAAGQAKOTSAYEFLP--PLSYRRCQGSFALVKNYATR 236

```
QY 116 IIMKEBEGEIKGOLISSADIAATVQHTMACHVKTTHRAILFCORDL--PQNNAV 173
DB 237 HIERESTLHAPDALPDYEAFCAPUKGVTGRHMLHRTORALFEYCEBKFLPSDAEPHS 296
QY 174 LVASGVASNYIRALIELTNATQCTLLCPPEPLCTONGIMIANGLERLAA--GLGL 231
DB 297 LVASGVACNDVIRNALLSSMAQGEYSYRPPKCLCTONGTMIANNGEKLAKDTAEMT 356
QY 232 HDIEGIRYEPKCPGLVDISKEVGEASIKVPLQKMEI 267
DB 357 TKYEQVDISGCKPIGDSDLIDVKEANLACKAKVADI 392

RESULT 11
Q96086 PRELIMINARY; PRT; 323 AA.
AC 096086;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE 1D37221P.
GN ORFNames=CG14231;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxId=7227;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Gurin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nuncio J., Pacled J., Paragas V., Park S., Phouenavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celinker S.;
RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY051882; AK93306.1; -
DR FlyBase; FBgn001060; CG14231.
DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009095; Peptidase M22.
DR InterPro; IPR009180; Pept M22_Osialgl.
DR Pfam; PF00814; Peptidase M22; 1.
DR PIRSF; PIRSF004537; Osialglc_pptds; 1.
DR PRINTS; PR00789; OSIALOPTASE.
SQ SEQUENCE 323 AA; 35828 MW; BBD8B54D9A2BF35A CRC64;

Query Match 30.0%; Score 415.5; DB 2; Length 323;
Best Local Similarity 38.2%; Pred. No. 2.7e-27;
Matches 100; Conservative 43; Mismatches 104; Indels 15; Gaps 6;

QY 1 MEAHALTIRLTN--KVEFPLVLLISGHCMLALVOQYSDFLLEKSIDIAPGMDLVYA 58
DB 50 MEAHALQARMEHPBQIGVPLCLLASGHCQLLVANGRGRLTLGQTLDDAPGEAFDKIG 109
QY 59 RLISLTHPEPCSTMSGKALIEHLAK--QGNRFHFDIKPPLHAKNCDSFPTGLQHTVTKII 117
DB 110 RLRLHLHLPEYRLNNGGRALIEHAQLASDPLAYEPFLPLAQORNCNFSFAGIKNNSFPAI 169
QY 118 MKKEKEGIEKGOILSSAADIATVQHTMACHVKTTHRAILFC--KQRDLIPQNNAV 175
DB 170 RABERARTRPPDGVISYVGDPCAGILSSVSRHMLHRTORALBYCLLPHRQLFGDTPPLV 229
QY 176 ASGGVANSFYIRALELTNATQCTLLCPPEPLCTONGIMIANGLERLGLIHDIE 235
DB 230 MSGGVANNDAIYANIEHLAAYGCRSPRPSKRYCSDNQVMIAMGVEQL-----LDKKE 283
QY 236 -GIRYEPKCPGLVDISKEVGEA 256
DB 284 ASTRYDYD--SIDIQSAGFA 302
```

```
RESULT 12
Q9VWD6 PRELIMINARY; PRT; 409 AA.
AC 09VWD6;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE CG14231-PA.
GN ORFNames=CG14231;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxId=7227;
RX [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazer J.R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabo G.L.,
RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benoe P.V., Berman B.P., Bhandari D., Bolhakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotier P.,
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durkin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA Jalili B., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mervilov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacled J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Pui V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wassenaar D.A., Weinstein G.M., Weisenbach J.,
RA Williams S.M., Woodgett W., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195 (2000).

[2]
RN
RP SEQUENCE FROM N.A.
RC MEDLINE=22426065; PubMed=12537568;
RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Laverly T., Muzny D.M., Nelson C.R.,
RA Pacled J.M., Park S., Pfeiffer B.D., Richards S., Sodergren B.J.,
RA Svirskas R., Taber P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun: Release 3 of the Drosophila
RL melanogaster euchromatic genome sequence."
RN Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).

[3]
RN
RP SEQUENCE FROM N.A.
RC MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J., Svirskas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
```

RA Ashburner M., Celniker S.E.;
 RT "The transposable elements of the Drosophila melanogaster euchromatin:
 a genomics perspective."
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426069; PubMed=12537572;
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
 RA Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 RA Lewis S.E.;
 RT "Annotation of the Drosophila melanogaster euchromatic genome: a
 RT systematic review";
 RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
 RN [5]
 RP SEQUENCE FROM N.A.
 RG FlyBase;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP SEQUENCE FROM N.A.
 RG FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AEO03513; AAF49008.1; -
 DR IntAct; Q9VMD6; -
 DR FlyBase; FBgn0031060; CG14231.
 DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.
 DR GO; GO:0008270; F:zinc ion binding; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR009180; Pept M22_Osialgl.
 DR Pfam; PF00814; Peptidase_M22; 1.
 DR PIRSF; PIRSF004537; Osialglc_pptds; 1.
 DR PRINTS; PR00789; OSIALOPTASE.
 DR ProDom; PD002367; Peptidase_M22; 1.
 DR TIGRFBMS; TIGR00329; gcp; 1.
 SQ SEQUENCE 409 AA; 45328 MW; 9797F667D155538 CRC64;
 Query Match 30.0%; Score 415.5; DB 2; Length 409;
 Best Local Similarity 38.2%; Pred. No. 3.6e-27;
 Matches 100; Conservative 43; Mismatches 104; Indels 15; Gaps 6;
 QY 1 MEAHALTRLTN--KVERPVLVILISGGHCLALVQGVSDFLILGKSLDIAPGMDLDRVA 58
 DB 136 MEALALQARMEHPBOIGYPLCLLASGGHCDLVANGPRLTLLOQTLDDAGAEAFDKIG 195
 QY 59 RLSLIKHPECSTWSGKAIEHLAK-QGNRFHFDIKPILHAKNCDFSPGLQHTVDKII 117
 DB 196 RLRLHLILPEYRLMNGGRAIEHAQALADPLAYEPFLPLAQRNCNFSFAGIKNNFRAI 255
 QY 118 MKKEKEGIEKGQILSSADIAATVQHTMACHLVKTRHAILFC--KORDLLPQNNAVLV 175
 DB 256 RARERARTRPPDGVITSNYGCACGLRSVSRHLMRTQRAIEYCLLPHQRLGDPPTLV 315
 QY 176 ASGGVANSFYRRALDELITNATQCTLLCPPRLCTDNGIMTANGIERKAGLGIIHDE 235
 DB 316 MSGGVANNDAIYANIEHLAAYGCRSFRPSKRYCSDNGVMIAHGEVQ-----LQDKK 369
 QY 236 -GIRYEPKPLGVNISKEVGEA 256
 DB 370 ASTRIYD--SIDIQSAGFA 388
 RESULT 13
 022145 PRELIMINARY; PRT; 480 AA.
 AC 022145; Q8VWL2;
 DT 01-JAN-1998 (TRENBLrel. 05, Created)
 DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)
 DT 05-JUL-2004 (TRENBLrel. 27, Last annotation update)
 DE Putative O-sialoglycoprotein endopeptidase (Sialoglycoprotease

DE GCPI).
 GN Name=At2g45270; Synonyms=GCPI;
 OS Arabidopsis thaliana (Mouse-ear cress).
 ON Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Rounsley S.D., Lin X., Ketchum K.A., Crosby M.L., Brandon R.C.,
 RA Sykes S.M., Mason T.M., Kerlavage A.R., Adams M.D., Somerville C.R.,
 RA Venter J.C.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Town C.D., Kaul S.;
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Hauswirth K., Adamska I.;
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
 RA Goldsmith A.D., Lee J.M., Quach H.L., Toriumi M., Yu G., Bowser L.,
 RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
 RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
 RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,
 RA Seki M., Shim P., Southwick A., Shinzaki K., Davis R.W., Ecker J.R.,
 RA Theologis A.;
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Banh J., Chan M.M., Chang C.H., Chang E., Dale J.M.,
 RA Deng J.M., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.L.,
 RA Tang C., Toriumi M., Wu H.C., Yamamura Y., Yu G., Bowser L.,
 RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
 RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
 RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,
 RA Seki M., Shim P., Southwick A., Shinzaki K., Davis R.W., Ecker J.R.,
 RA Theologis A.;
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; ACO02387; AAB82636.2; -
 DR EMBL; AY024338; AAK00530.1; -
 DR EMBL; AY063864; AAL36220.1; -
 DR EMBL; AY117283; AAM51358.1; -
 DR PIR; E84888; E84888.
 DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.
 DR GO; GO:0008233; F:peptidase activity; IEA.
 DR GO; GO:0008270; F:zinc ion binding; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR009095; Peptidase_M22.
 DR InterPro; IPR009180; Pept M22_Osialgl.
 DR Pfam; PF00814; Peptidase_M22; 1.
 DR PIRSF; PIRSF004537; Osialglc_pptds; 1.
 DR PRINTS; PR00789; OSIALOPTASE.
 DR ProDom; PD002367; Peptidase_M22; 1.
 DR TIGRFBMS; TIGR00329; gcp; 1.
 KW Protease.
 SQ SEQUENCE 480 AA; 52995 MW; 20DD6A86ACCF1FPAD CRC64;
 Query Match 29.4%; Score 407.5; DB 2; Length 480;
 Best Local Similarity 37.0%; Pred. No. 2.1e-26;
 Matches 104; Conservative 37; Mismatches 93; Indels 47; Gaps 7;
 QY 1 MEAHALTRLTN-KVERPVLVILISGGHCLALVQGVSDFLILGKSLDIAPGMDLDRVA 59
 DB 195 MEHALVALVREBELSFFPMALLISGGHNLVLAHKGGYTGQTLVDDAIGAEAFDTAK 254
 QY 60 RLSLIKHPECSTWSGKAIEHLAKQGNRFHFDIKPILHAKNCDFSPGLQHTVDKIIK 119
 DB 255 WGLGLDMH-----RSGPAVEALLEGDAKSVKENVVPMKHKDCNFSYAGLKTQVRLAIEA 309

```
QY 120 KEKEGIEKQILSSA-----ADINATQHTMACHLYKTRHAILFCCKORDLLPQNN 171
DB 310 KE-----IDAKCPVSSATNEDRRNRADIAASFQRAVAILERKCEBAIIMALE---LEPSI 362
QY 172 AVIVASGVASNFYIRRALIELITNATQCTLLCPRLCTDNGIMIAMNGIERLAGLGL 231
DB 363 KHVIVSGGVASNKYVRLRNNIVENKNIKLVCPPLSLCTDNGIMIAMNGIERLAGLGL 418
QY 232 HDIEGIRYE-----PKCPLGVDSKEVGEA 256
DB 419 -----RYDPPPPATEPEPDYVYDRLPRRPLGEBAVAKGSEA 453

RESULT 14
Q73H71 PRELIMINARY; PRT; 335 AA.
ID Q73H71
AC Q73H71
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE Peptidase, M22 family protein.
OS OrderelocutNames=MD0699;
OC Wolbachia pipiensis wMel.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Wolbachiae; Wolbachia.
OX NCBI_TaxID=66077;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=15024419;
RA Wu M., Sun L.V., Vamathevan J.J., Riegler M., Deboy R.T.,
RA Brownlie J.C., McGraw E.A., Martin W., Esser C., Ahmadijad N.,
RA Miegand C., Madupu R., Beaman M.J., Brinkac L.M., Daugherty S.C.,
RA Durkin A.S., Kolonay J.F., Nelson W.C., Mohamoud Y., Lee P.,
RA Berry K.J., Young M.B., Utterback T.R., Weidman J.P., Nierman W.C.,
RA Paulsen I.T., Nelson K.B., Tettelin H., O'Neill S.L., Eisen J.A.,
RT "Phylogenomics of the reproductive parasite Wolbachia pipiensis wMel:
RT a streamlined genome overruled by mobile genetic elements."
RL PLOS Biol. 2:327-341(2004).
DR EMBL; AE017258; AAS14395.1;
DR TIGR; WD0699;
DR GO; GO:0008450; P:O-sialoglycoprotein endopeptidase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000905; Peptidase M22.
DR InterPro; IPR009180; Pept M22_Osialgl.
DR Pfam; PF00814; Peptidase_M22; 1.
DR PIRSF; PIRSF004537; Osialglc_ptds; 1.
DR PRINTS; PR00789; OSIALOPTASE.
DR ProDom; PD002367; Peptidase_M22; 1.
DR TIGRFAMs; TIGRFAMs; gcp; 1.
DR Complete proteome.
SQ SEQUENCE 335 AA; 36634 MW; F4914CDA58BC9390 CRC64;

Query Match 26.6%; Score 369; DB 2; Length 335;
Best Local Similarity 34.7%; Pred. No. 2.8e-23;
Matches 84; Conservative 46; Mismatches 86; Indels 26; Gaps 5;

QY 1 MEHAALTRLTNKKVEPPFLVLLISGGHCLALVGVSDPLLGKSLDIAPGMDLDKVAR 60
DB 112 LEHALVTRLTNKKVEPPFLVLLISGGHCLALVGVSDPLLGKSLDIAPGMDLDKVAR 171
QY 61 LSLIKHECSTSGSKALEHLAKQGNRFHPIKPLHAKNCDFSTGLQHTVDKIINK 120
DB 172 LGL-----SYGSPPLIEKAKKNGTRFKLPRAMIKRSGCNFSFGSKTRAVKNLVDEL 224
QY 121 EKEBGIRKQILSSAADIATVQHTMACHLYKTRHAILFCCKORDLLPQNNAVIVASGV 180
DB 225 KMSF-----QDVCVGASFQECISIDILDRVSNALIMBSALIKIND---FYITGCV 273
QY 191 ASNFYIRRALIELITNATQCTLLCPRLCTDNGIMIAMNGIERLAGLGL 240
DB 274 AANFVIRALIELITNATQCTLLCPRLCTDNGIMIAMNGIERLAGLGL 325
```

```
QY 241 PK 242
DB 326 PR 327

RESULT 15
Q92LH8 PRELIMINARY; PRT; 360 AA.
ID Q92LH8
AC Q92LH8;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE PROBABLE O-SIALOGLYCOPROTEIN ENDOPEPTIDASE (EC 3.4.24.57).
OS ORFNames=Smc03230;
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=1021;
RC MEDLINE=21396507; PubMed=11481430; DOI=10.1073/pnas.161294398;
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA Boiteard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Godrie T., Goffeau A., Kahn D., Kiss B., Lelaure V., Masny D.,
RA Pohl T., Portetalle D., Puehler A., Purnelle B., Ramsperger U.,
RA Renard C., Thebaud P., Vandenbol M., Weidner S., Galibert F.;
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021."
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR EMBL; AL591792; CAC47657.1;
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000905; Peptidase_M22.
DR InterPro; IPR009180; Pept M22_Osialgl.
DR Pfam; PF00814; Peptidase_M22; 1.
DR PIRSF; PIRSF004537; Osialglc_ptds; 1.
DR PRINTS; PR00789; OSIALOPTASE.
DR ProDom; PD002367; Peptidase_M22; 1.
DR TIGRFAMs; TIGRFAMs; gcp; 1.
DR Complete proteome; Hydrolase.
SQ SEQUENCE 360 AA; 37906 MW; A07F946AB562EAB6 CRC64;

Query Match 25.7%; Score 356.5; DB 2; Length 360;
Best Local Similarity 35.6%; Pred. No. 3.5e-22;
Matches 95; Conservative 38; Mismatches 91; Indels 43; Gaps 7;

QY 1 MEHAALTRLTNKKVEPPFLVLLISGGHCLALVGVSDPLLGKSLDIAPGMDLDKVAR 60
DB 116 LEHALVTRLTNKKVEPPFLVLLISGGHCLALVGVSDPLLGKSLDIAPGMDLDKVAR 175
QY 61 LSLIKHECSTSGSKALEHLAKQGNRFHPIKPLHAKNCDFSTGLQHTVDKIINK 120
DB 176 LGL-FYP-----GGPVRERAQAGNAERFDPRLVGPARDLFSFGSKTRAVQQAQSL 228
QY 111 -HTVDKIINKKEKEBGIRKQILSSAADIATVQHTMACHLYKTRHAILFCCKORDLLPQ 169
DB 229 GAVTDQDI-----ADVCAFSFORAISRTILDRVGRGKRFR-ADFAV 269
QY 170 NNAVIVASGVASNFYIRRALIELITNATQCTLLCPRLCTDNGIMIAMNGIERLAGL 229
DB 270 DQPALVIVASGVANQTLRRTLOSICDEHGFRTIAPLQICTDPAAMIAAGABRLLAGL- 328
QY 230 ILHDIEGIRYEPKC--PLGVDSKEVG 254
DB 329 ---PADGIDAPRSRMPDSEAKALIG 352

Search completed: November 10, 2005, 16:52:02
Job time : 65.7185 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using SW model

Run on: November 10, 2005, 15:51:52 ; Search time 21.4761 Seconds
(without alignments)
928.069 Million cell updates/sec

Title: US-10-649-273-2_COPY_148_414

Perfect score: 1385
Sequence: 1 MEAHLATRLTNKVEPPLV.....DISKVGESAIKVPOLKMEI 267

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA:*

- 1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/1/1aa/PCTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/1/1aa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1385	100.0	267	US-10-067-443-22	Sequence 22, Appl
2	1385	100.0	414	US-10-067-443-2	Sequence 2, Appl
3	1358.5	98.1	439	US-10-067-443-19	Sequence 19, Appl
4	409.5	29.6	463	US-10-067-443-3	Sequence 3, Appl
5	307	22.2	350	US-09-540-226-2726	Sequence 2726, Ap
6	291.5	21.0	401	US-09-252-991A-17372	Sequence 17372, A
7	280.5	20.3	357	US-09-543-681A-6513	Sequence 6513, Ap
8	279	20.1	342	US-08-087-797-3	Sequence 3, Appl
9	276	19.9	421	US-10-067-443-4	Sequence 4, Appl
10	275	19.9	421	US-10-067-443-28	Sequence 28, Appl
11	270.5	19.5	325	US-08-087-797-2	Sequence 2, Appl
12	268.5	19.4	343	US-09-489-039A-9221	Sequence 9221, Ap
13	261	18.8	363	US-09-107-532A-6609	Sequence 6609, Ap
14	254	18.3	336	US-08-987-121A-4	Sequence 4, Appl
15	250	18.1	335	US-08-961-083-52	Sequence 52, Appl
16	250	18.1	335	US-09-536-784-52	Sequence 52, Appl
17	248	17.9	336	US-09-066-512-2	Sequence 52, Appl
18	246	17.8	336	US-09-583-110-4857	Sequence 4857, Ap
19	243	17.5	327	US-10-067-443-5	Sequence 5, Appl
20	241	17.4	336	US-09-107-433-4221	Sequence 4221, Ap
21	236	17.0	273	US-09-710-279-728	Sequence 728, App
22	236	17.0	366	US-09-134-000C-4956	Sequence 4956, Ap
23	233	16.8	368	US-09-134-001C-3909	Sequence 3909, Ap
24	220	15.5	341	US-09-149-624-2	Sequence 2, Appl
25	218	15.7	344	US-09-198-452A-213	Sequence 213, App
26	218	15.7	360	US-09-438-185A-196	Sequence 196, App
27	211.5	15.3	344	US-09-602-777A-148	Sequence 148, App

28	194	14.0	340	US-10-067-443-6	Sequence 6, Appl
29	181.5	13.1	143	US-09-328-352-4387	Sequence 4387, Ap
30	169	12.2	292	US-09-724-623-81	Sequence 81, Appl
31	157	11.3	214	US-09-328-352-4609	Sequence 4609, Ap
32	94.5	6.8	480	US-09-583-110-5050	Sequence 5050, Ap
33	94.5	6.8	481	US-09-107-433-3197	Sequence 3197, Ap
34	83	6.0	42	US-09-902-540-13841	Sequence 13841, A
35	81.5	5.9	1087	US-08-264-002-5	Sequence 5, Appl
36	81	5.8	1072	US-09-902-540-15572	Sequence 15572, A
37	79	5.7	328	US-09-710-279-3008	Sequence 3008, Ap
38	79	5.7	336	US-09-134-001C-5549	Sequence 5549, Ap
39	79	5.7	445	US-08-083-945C-2	Sequence 2, Appl
40	79	5.7	445	US-08-083-945C-7	Sequence 7, Appl
41	79	5.7	578	US-09-252-991A-31318	Sequence 31318, A
42	79	5.7	1137	US-09-538-092-968	Sequence 968, App
43	78.5	5.7	503	US-09-252-991A-22790	Sequence 22790, A
44	78.5	5.7	1658	US-08-609-0949A-13	Sequence 13, Appl
45	78.5	5.7	1658	US-09-170-996-13	Sequence 13, Appl

ALIGNMENTS

RESULT 1		US-10-067-443-22		Sequence 22, Application US/10067443	
				Patent No. 6642041	
				GENERAL INFORMATION:	
				APPLICANT: Bristol-Myers Squibb Company	
				TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED I	
				FILE REFERENCE: D0073 NP	
				CURRENT APPLICATION NUMBER: US/10/067,443	
				CURRENT FILING DATE: 2002-02-05	
				PRIOR APPLICATION NUMBER: US 60/266, 518	
				PRIOR FILING DATE: 2001-02-05	
				PRIOR APPLICATION NUMBER: US 60/282, 814	
				PRIOR FILING DATE: 2001-04-10	
				NUMBER OF SEQ ID NOS: 71	
				SOFTWARE: Patentin version 3.0	
				SEQ ID NO 22	
				LENGTH: 267	
				TYPE: PRT	
				ORGANISM: homo sapiens	
				US-10-067-443-22	
				Query Match	
				Best Local Similarity 100.0%; Score 1385; DB 4; Length 267;	
				Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1	MEAH	ATRLTNKVEPPLVLLISGHCILAAVGVSPFLIGKSLDINPGLMVKVAR	60	
DB	1	MEAH	ATRLTNKVEPPLVLLISGHCILAAVGVSPFLIGKSLDINPGLMVKVAR	60	
QY	61	LSL	HPESCTSGGKATIEHLAKOGRPFHDIKPLPHAKNCDPFTGQHYTDKIMK	120	
DB	61	LSL	HPESCTSGGKATIEHLAKOGRPFHDIKPLPHAKNCDPFTGQHYTDKIMK	120	
QY	121	EKEE	BIEKQILSSAADIAATVOHTMACHLVKTRRAILFCRKORDILPNNNAVLSAGCV	180	
DB	121	EKEE	BIEKQILSSAADIAATVOHTMACHLVKTRRAILFCRKORDILPNNNAVLSAGCV	180	
QY	181	ASN	PIRRLATLTNAOCTLLCPPLRLCTDNGIMIANWGIRLRAGLILHDIGIRYE	240	
DB	181	ASN	PIRRLATLTNAOCTLLCPPLRLCTDNGIMIANWGIRLRAGLILHDIGIRYE	240	
QY	241	PKCP	GVDSKVGESAIKVPOLKMEI	267	
DB	241	PKCP	GVDSKVGESAIKVPOLKMEI	267	
RESULT 2		US-10-067-443-2			

```
; Sequence 2, Application US/10067443
; Patent No. 6642041
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE HIGHLY EXPRESSED IN
; FILE OF INVENTION: SPINAL CORD, MP-1
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 2
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-067-443-2
```

```
Query Match          100.0%; Score 1385; DB 4; Length 414;
Best Local Similarity 100.0%; Pred. No. 7.3e-160;
Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 MEAHLTRLTNKKVEPFLVLLISGHCILALVQGVSDPFLLGKSLDIAPGMDLKVARR 60
Db 148 MEAHLTRLTNKKVEPFLVLLISGHCILALVQGVSDPFLLGKSLDIAPGMDLKVARR 207
Qy 61 LSLIKHPECSTMSGKAIETHLAKQGNRFHDIKPLHAKNCDFSFTGLQHTVDKTIIMKK 120
Db 208 LSLIKHPECSTMSGKAIETHLAKQGNRFHDIKPLHAKNCDFSFTGLQHTVDKTIIMKK 267
Qy 121 EKEGIEKGQILSSAADIATVQHTMACHLVKRTIRALIFCKQRDLPPONNAVIVASGCV 180
Db 268 EKEGIEKGQILSSAADIATVQHTMACHLVKRTIRALIFCKQRDLPPONNAVIVASGCV 327
Qy 181 ASNFYIRRALIELTNATQCTLLCPPRLCTDNGIMIANNGIERLRAGILHDIGIRY 240
Db 328 ASNFYIRRALIELTNATQCTLLCPPRLCTDNGIMIANNGIERLRAGILHDIGIRY 387
Qy 241 PKCPLGVDISKEVGEASIKVPOLKMEI 267
Db 388 PKCPLGVDISKEVGEASIKVPOLKMEI 414
```

```
RESULT 3
US-10-067-443-19
; Sequence 19, Application US/10067443
; Patent No. 6642041
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE HIGHLY EXPRESSED IN
; FILE OF INVENTION: SPINAL CORD, MP-1
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 19
; LENGTH: 439
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-067-443-19
```

```
Query Match          98.1%; Score 1358.5; DB 4; Length 439;
Best Local Similarity 91.1%; Pred. No. 1.4e-156;
Matches 266; Conservative 0; Mismatches 1; Indels 25; Gaps 1;
```

```
Qy 1 MEAHLTRLTNKKVEPFLVLLISGHCILALVQGVSDPFLLGKSLDIAPGMDLKVARR 60
Db 148 MEAHLTRLTNKKVEPFLVLLISGHCILALVQGVSDPFLLGKSLDIAPGMDLKVARR 207
Qy 61 LSLIKHPECSTMSGKAIETHLAKQGNRFHDIKPLHAKNCDFSFTGLQHTVDKTIIMKK 120
Db 208 LSLIKHPECSTMSGKAIETHLAKQGNRFHDIKPLHAKNCDFSFTGLQHTVDKTIIMKK 267
Qy 121 EKEGIEKGQILSSAADIATVQHTMACHLVKRTIRALIFCKQRDLPPONNAVIVASGCV 180
Db 268 EKEGIEKGQILSSAADIATVQHTMACHLVKRTIRALIFCKQRDLPPONNAVIVASGCV 327
Qy 181 ASNFYIRRALIELTNATQCTLLCPPRLCTDNGIMIANNGIERLRAGILHDIGIRY 240
Db 328 ASNFYIRRALIELTNATQCTLLCPPRLCTDNGIMIANNGIERLRAGILHDIGIRY 387
Qy 241 PKCPLGVDISKEVGEASIKVPOLKMEI 267
Db 388 IAMNGIERLRAGILHDIGIRYEPKCPPLGVDISKEVGEASIKVPOLKMEI 439
```

```
RESULT 4
US-10-067-443-3
; Sequence 3, Application US/10067443
; Patent No. 6642041
```

```
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE HIGHLY EXPRESSED IN
; FILE OF INVENTION: SPINAL CORD, MP-1
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 3
; LENGTH: 463
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-067-443-3
```

```
Query Match          29.6%; Score 409.5; DB 4; Length 463;
Best Local Similarity 36.6%; Pred. No. 8.2e-41;
Matches 100; Conservative 37; Mismatches 91; Indels 45; Gaps 6;
```

```
Qy 1 MEAHLTRLTNKKVEPFLVLLISGHCILALVQGVSDPFLLGKSLDIAPGMDLKVARR 59
Db 192 MEAHLTRLTNKKVEPFLVLLISGHCILALVQGVSDPFLLGKSLDIAPGMDLKVARR 251
Qy 60 RSLIKHPECSTMSGKAIETHLAKQGNRFHDIKPLHAKNCDFSFTGLQHTVDKTIIMKK 119
Db 252 WGLDMH-----RSGGPAVEBELLEGDAKSVKFNVMKHKCNFSYAGIKCTQVRLAIEA 306
Qy 120 EKEGIEKGQILSSAADIATVQHTMACHLVKRTIRALIFCKQRDLPPONNAVIVASGCV 179
Db 307 KE-----IRNADIAASFQVAVLHLEKRCRAIDWALE---LEPSIKHMYISGG 353
Qy 180 VASNFYIRRALIELTNATQCTLLCPPRLCTDNGIMIANNGIERLRAGILHDIGIRY 239
Db 354 VASNKYVRLRNNIVANKKILVCPPLSLCTDNGVAVNTGLEHFRVG-----RX 403
Qy 240 E-----PKCPLGVDISKEVGEA 256
Db 404 DPPPATPEDEYVYDLRPRWPLGEEYAKGRSEA 436
```

```
RESULT 5
US-09-540-236-2726
; Sequence 2726, Application US/09540236
; Patent No. 6673910
```


GENERAL INFORMATION:
APPLICANT: Gary H. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATAR
FILE REFERENCE: 2709.2005-001
CURRENT APPLICATION NUMBER: US/09/540.236
NUMBER OF SEQ ID NOS: 3840
SEQ ID NO 2726
LENGTH: 350
TYPE: PRT
ORGANISM: M. catarrhalis
US-09-540-236-2726

Query Match 22.2%; Score 307; DB 4; Length 350;
Best Local Similarity 33.3%; Pred. No. 1.7e-28;
Matches 84; Conservative 37; Mismatches 91; Indels 40; Gaps 7;

QY 1 MEAHATIRLT--NKVEPFLVLLISGHCILALVQVSDFLILGKSLDIAPGDMLDKVA 58
DB 119 MEGHILAPLSDPSPFPVCLLVSGHTLVADGAVQVQILGESIDDAVGECDKTA 178
QY 59 RLSLIKHPECSTMSGGKAIEMHAKOGRFHFDIKRPLHAKKDCSFSTGLQHTVDKIM 118
DB 179 KMLKL-PYP-----GGPNIEKLAKNPAPAYELPRPQHR-KGLDFSSGKTAIHNLIK 230
QY 119 KEKEGIEKQILSSAADIAATVQHTMACHLVKTRTHALIFCQKRDLPQNNAVLVASG 178
DB 231 DTPMAQSDP-----ARRADIASFEYAVDTLVKCKTKALQMTGIRQ-----LVVAG 277
QY 179 GVAASFYIRALIELTNATQCTLLCPPELCTDNGIMIAMNGIERLRAG-----227
DB 278 GVSANQTLRRTLTETLRQIDASVYVAPTELCTDNGAMIAVAGFRLSRGQSDDLAVRCIP 337
QY 228 -----LGLIHD 233
DB 338 RWDMTMLGIEYD 349

RESULT 6

US-09-252-991A-17372
Sequence 17372, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252.991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074.788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094.190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 17372
LENGTH: 401
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17372

Query Match 21.0%; Score 291.5; DB 4; Length 401;
Best Local Similarity 35.4%; Pred. No. 1.6e-26;
Matches 87; Conservative 37; Mismatches 99; Indels 23; Gaps 9;

QY 1 MEAHATIRLTNK-VEPFLVLLISGHCILALVQVSDFLILGKSLDIAPGDMLDKVAR 59
DB 172 MEGHILAPLSEPPFPFVALVSGHTLVADGAVQVQILGESIDDAVGEAPDKTAK 231
QY 69 RLSLIKHPECSTMSGGKAIEMHAKOGRFHFDIKRPLHAKKDCSFSTGLQHTVDKIM 119
DB 232 LIGL-GYP-----GGPEIRALBERGTGRFVPRPMTDRPGDFSGKTKPTLTLN-TWQ 283

QY 120 KEKEGIEKQILSSAADIAATVQHTMACHLVKTRTHALIFCQKRDLPQNNAVLVASG 179
DB 284 RCVEAGDDSEQ---TRCDIALAFQTAIVETLLIKCRRL---KOTGL--KN---LVVAG 332
QY 180 VASNFYIRALIELTNATQCTLLCPPELCTDNGIMIAMNGIERLRAGLGIHDIIGIRY 239
DB 333 VSANQALSGLEKMEKGMQVYVAPRPTDNGAMIAVAGCQRLLAG---QHDGPAISV 389
QY 240 EPKCPL 245
DB 390 QPRWPM 395

RESULT 7

US-09-543-681A-6513
Sequence 6513, Application US/09543681A
Patent No. 6605709
GENERAL INFORMATION:
APPLICANT: GARY BRETON
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
FILE REFERENCE: 2709.1002-001
CURRENT APPLICATION NUMBER: US/09/543.681A
CURRENT FILING DATE: 2000-04-05
PRIOR APPLICATION NUMBER: US 60/128.706
PRIOR FILING DATE: 1999-04-09
NUMBER OF SEQ ID NOS: 8344
SEQ ID NO 6513
LENGTH: 357
TYPE: PRT
ORGANISM: Proteus mirabilis
US-09-543-681A-6513

Query Match 20.3%; Score 280.5; DB 4; Length 357;
Best Local Similarity 32.7%; Pred. No. 3e-25;
Matches 81; Conservative 35; Mismatches 101; Indels 31; Gaps 8;

QY 1 MEAHATIRLTNKV-EPPFLVLLISGHCILALVQVSDFLILGKSLDIAPGDMLDKVAR 59
DB 129 MEGHILAPMLEKTPDFPFVALVSGHTQLISVIGISYTLIGESIDDAVGEAPDKTAK 188
QY 60 RLSLIKHPECSTMSGGKAIEMHAKOGRFHFDIKRPLHAKKDCSFSTGLQHTVDKIM 119
DB 189 LIGL-DYP-----GGPVLSSKWAQGVGGRFVPRPMTDRPGDFSGKTKPTAANTIQ 241
QY 120 KEKEGIEKQILSSAADIAATVQHTMACHLVKTRTHALIFCQKRDLPQNNAVLVASG 178
DB 242 NDDSE-----QTRADIAFAFEDAVDTLAIKCRRA-----LEQTGFKRLVMAG 284
QY 179 GVAASFYIRALIELTNATQCTLLCPPELCTDNGIMIAMNGIERLRAGL-GIHDIEGI 237
DB 285 GVSANRTLRKAKMIMBOLGGEVYVAPRPTDNGAMIALAGMIRKGGTEGPL---GV 340
QY 238 RYBPKCPL 245
DB 341 TVRPRWPL 348

RESULT 8

US-08-087-797-3
Sequence 3, Application US/08087797
Patent No. 5543312
GENERAL INFORMATION:
APPLICANT: Mellors, Alan
APPLICANT: Lo, Reggie Y.C.
APPLICANT: Abdullah, Khalid M.
TITLE OF INVENTION: Pasteurella Haemolytica
TITLE OF INVENTION: Glycoprotease
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSER: Bell, Seltzer, Park & Gibson, P.A.
STREET: 1211 East Morehead Street,

```

; CITY: Charlotte
; STATE: No. 5543312th Carolina
; COUNTRY: United States
; ZIP: 28234
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/087,797
; FILING DATE: 14-JUL-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Layton, Jr., Samuel G.
; REGISTRATION NUMBER: 22807
; REFERENCE/DOCKET NUMBER: 3374-80
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 704 377 1561
; TELEFAX: 704 334 2014
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 342 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-087-797-3

```

Query Match 20.1%; Score 279; DB 1; Length 342;

Best Local Similarity 33.6%; Pred. No. 4.2e-25;

Matches 83; Conservative 32; Mismatches 108; Indels 24; Gaps 8;

```

QY 1 MEAAHLLTIRLT-TNVEFPFVLVLLISGGHCLLALVGVSDPFLLGKSLDIAPGMDLKVAR 59
DB 112 MEGHLLPMDNDPPEFFVALVSGHQLISVGLQVELGSEIDPAAGEADPKTAK 171
QY 60 RLSTIKHPECSTMSGKAIIEHLAKQGNRFHFDIKPPLHAKNCDPSTFTGLQHYTDKI 119
DB 172 LGLG-DYF-----GGPLSTGMAAGTIGRVRFPFPMTRDRLDSFSGLKTFPAANTIRD 224
QY 120 KEKEEGIEKGQILSSAADIAATVQHTMACHLVKTRTHAILFCQKRDLLPQNN-AVLVAVSG 178
DB 225 NXXXXXGXTDDQ---TRADIAFAFEDAVVDITMIKKRA-----LDQTFKKLIVMAG 272
QY 179 GVAENFYIRRALBITLTAATCTLLCPPRRLCTDNGIMTANNGIERLRAGLGIHDIIEGR 238
DB 273 GVSANRTRLRAKLAEMMKKRGVEVFYARPEFCTDNGAMTAYAGMVFKA--GATADL-GVS 329
QY 239 YEPKCP 245
DB 330 VRPRWPL 336

```

RESULT 9

US-10-067-443-4

```

; Sequence 4, Application US/10067443
; Patent No. 6642041
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED I
; TITLE OF INVENTION: SPINAL CORD, MP-1
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 4
; LENGTH: 421

```

```

; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
; US-10-067-443-4

```

Query Match 19.9%; Score 276; DB 4; Length 421;

Best Local Similarity 30.4%; Pred. No. 1.4e-24;

Matches 85; Conservative 52; Mismatches 107; Indels 36; Gaps 8;

```

QY 1 MEAAHLLTIRLT-NKVEFPFVLVLLISGGHCLLALVGVSDPFLLGKSLDIAPGMDLKVAR 59
DB 134 MRAHALLSTLVDSVRFPSAVLLISGHALISVADEVKFKLYGQSVSSPGECIDKVAR 193
QY 60 RLSTIKHPECSTMSGKAIIEHLAKQGN--RFHFDIKPPLHAKNCDPSTFTGLQHYTDKI 116
DB 194 QLGDL-GSEFPDGIHGAAVEIILASRASADGHLRYPIFLPNVPRKANNFPQIKSGSYLTLLE 252
QY 117 IMKKEEGIEKGQILSSAADIAATVQHTMACHLVKTRTHAILFCQKRDLLPQNNAVLVA 176
DB 253 RLKNSETSID-----IPDFCASLQNTVAHHSKLIHFESLSQEKLPQO---LVI 302
QY 177 SGVANSFYIRRALBITLTAATCTLLCPPRRLCTDNGIMTANNGIERLRAGLGIHDIIEGR 236
DB 303 GGVANAOYIFGALISKLSAAHNVTTIKVLSTCTDNAMEMAYSGL-----LMLVNRSEA 356
QY 237 IRYEP-----KCPLGVDISKVEGASIKVPQLM 265
DB 357 IWRPNIDIPDTIYAHARSDIGTDASSEI----IDTPRRKL 392

```

RESULT 10

US-10-067-443-28

Sequence 28, Application US/10067443

Patent No. 6642041

```

; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED I
; TITLE OF INVENTION: SPINAL CORD, MP-1
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 28
; LENGTH: 421
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
; US-10-067-443-28

```

Query Match 19.9%; Score 276; DB 4; Length 421;

Best Local Similarity 30.4%; Pred. No. 1.4e-24;

Matches 85; Conservative 52; Mismatches 107; Indels 36; Gaps 8;

```

QY 1 MEAAHLLTIRLT-NKVEFPFVLVLLISGGHCLLALVGVSDPFLLGKSLDIAPGMDLKVAR 59
DB 134 MRAHALLSTLVDSVRFPSAVLLISGHALISVADEVKFKLYGQSVSSPGECIDKVAR 193
QY 60 RLSTIKHPECSTMSGKAIIEHLAKQGN--RFHFDIKPPLHAKNCDPSTFTGLQHYTDKI 116
DB 194 QLGDL-GSEFPDGIHGAAVEIILASRASADGHLRYPIFLPNVPRKANNFPQIKSGSYLTLLE 252
QY 117 IMKKEEGIEKGQILSSAADIAATVQHTMACHLVKTRTHAILFCQKRDLLPQNNAVLVA 176
DB 253 RLKNSETSID-----IPDFCASLQNTVAHHSKLIHFESLSQEKLPQO---LVI 302
QY 177 SGVANSFYIRRALBITLTAATCTLLCPPRRLCTDNGIMTANNGIERLRAGLGIHDIIEGR 236
DB 303 GGVANAOYIFGALISKLSAAHNVTTIKVLSTCTDNAMEMAYSGL-----LMLVNRSEA 356
QY 237 IRYEP-----KCPLGVDISKVEGASIKVPQLM 265

```

Db 357 IMWRPNDIPTIYAHASDIDGTDSSEI-----IDTPRKL 392

RESULT 11

US-08-087-797-2

Sequence 2, Application US/08087797

Patent No. 5543312

GENERAL INFORMATION:

APPLICANT: Mellors, Alan

APPLICANT: Y. C.

APPLICANT: Abdullah, Khalid M.

TITLE OF INVENTION: Pasteurella Haemolytica

TITLE OF INVENTION: Glycopolase

TITLE OF INVENTION: Gene and the Purified Enzyme

NUMBER OF SEQUENCES: 3

CORRESPONDENCE ADDRESS:

ADDRESSEE: Bell, Seltzer, Park & Gibson, P.A.

STREET: 1211 East Morehead Street,

CITY: Charlotte

STATE: No. 5543312th Carolina

COUNTRY: United States

ZIP: 28234

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/087,797

FILING DATE: 14-JUL-1993

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Layton, Jr., Samuel G.

REGISTRATION NUMBER: 22807

REFERENCE/DOCKET NUMBER: 3374-80

TELECOMMUNICATION INFORMATION:

TELEPHONE: 704 377 1561

TELEFAX: 704 334 2014

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 325 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-087-797-2

Query Match 19.5%; Score 270.5; DB 1; Length 325;
Best Local Similarity 35.0%; Pred. No. 4.2e-24;
Matches 79; Conservative 26; Mismatches 102; Indels 19; Gaps 6;

QY 1 MEAHLATIRL-TNKVEPFLVLLISGHCILALVQVSDFLLGKSLDIAPGMDLKVAR 59
Db 112 MEGLHAPMLEBNAPPEPVALISGHTQLVKVDVQGYELLGSEIDDAAGEAFDXTKX 171
QY 60 RLSLIHPECSTMSGCAIEHLAKQGNRFHFDIKPPLHAKNCDPSFTGLQHTDKIIM 119
Db 172 LGLL-DYP-----AGVMSKLAESGTPNRPKPPMTDRPGLDFSPSGLKTFAANTTKA 224
QY 120 KEKEEGIEKQILSSADIAATVQHTMACHLVKRTHAILFCRKORDLLPQNNAVLVASCG 179
Db 225 NUNENGEIDBQ---TKCDIAHAQAV-----VDITILKCK-KALEGQYKRLVWAGG 273
QY 180 VASNFYIRALAILTNATQCTLLCPPRLCTDNGIMIAMNGIEBLR 225
Db 274 VSANKQIRADLAEMMKLKGSVFYPRPQFCTDNGAMIAVGYPLTKX 319

RESULT 12
US-03-403-039A-9221
Sequence 9221, Application US/09489039A
Patent No. 6610816
GENERAL INFORMATION:

APPLICANT: Gary Brelton et. al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
FILE REFERENCE: 2709.2004001
CURRENT APPLICATION NUMBER: US/09/489,039A
CURRENT FILING DATE: 2000-01-27
PRIOR APPLICATION NUMBER: US 60/117,747
PRIOR FILING DATE: 1999-01-29
NUMBER OF SEQ ID NOS: 14342
LENGTH: 343
TYPE: PRT
ORGANISM: Klebsiella pneumoniae
US-09-489-039A-9221

Query Match 19.4%; Score 268.5; DB 4; Length 343;
Best Local Similarity 32.7%; Pred. No. 8.1e-24;
Matches 81; Conservative 31; Mismatches 105; Indels 31; Gaps 8;

QY 1 MEAHLATIRL-TNKVEPFLVLLISGHCILALVQVSDFLLGKSLDIAPGMDLKVAR 59
Db 118 MEGLHAPMLEBNAPPEPVALISGHTQLISVIGGYELLGSEIDDAAGEAFDXTKX 177
QY 60 RLSLIHPECSTMSGCAIEHLAKQGNRFHFDIKPPLHAKNCDPSFTGLQHTDKIIM 119
Db 178 LGLL-DYP-----GGPMLSKWASQGTGGRFVFPKPPMTDRPGLDFSPSGLKTFAANTIRS 230
QY 120 KEKEEGIEKQILSSADIAATVQHTMACHLVKRTHAILFCRKORDLLPQNN-ATLVASG 178
Db 231 NGDDE-----QTRADIAAFEDAVVDTLMIKCRA-----LEQGFKRLVWAG 273
QY 179 GVSANFYIRALAILTNATQCTLLCPPRLCTDNGIMIAMNGIEBLRAGLIIHDI-GI 237
Db 274 GVSANRTIRAKLAEMQGRGGEVIFYARPEFCTDNGAMIAVGMVLTQGA-----KALEGV 329
QY 238 RYEPKPL 245
Db 330 TVRPWPL 337

RESULT 13
US-09-107-532A-6609
Sequence 6609, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02154
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Arinello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012

```

/
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (781)893-5007
/ TELEFAX: (781)893-8277
/ INFORMATION FOR SEQ ID NO: 6609:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 363 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ HYPOTHEICAL: YES
/ ORIGINAL SOURCE:
/ ORGANISM: Enterococcus faecium
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (B) LOCATION 1..363
/ SEQUENCE DESCRIPTION: SEQ ID NO: 6609:
/
/ US-09-107-532A-6609
/
/ Query Match 18.8% Score 261; DB 4; Length 363;
/ Best Local Similarity 33.8% Pred No. 7.3e-23;
/ Matches 75; Conservative 37; Mismatches 78; Indels 32; Gaps 9
/
/ Qy 1 MEAHATRIITNKVPEPFLVLLISGSHCLAVGVSDPFLGSLDIAPGMDLKVARR 60
/ Db 140 MAGHIYARLVKVPFQPLMALLVSGHTELVMQEDSGYIEIGETRDDAAGAYKVGKRV 199
/
/ Qy 61 LSLIKHEPCSTMSGKALIEHLAKQ-NRPHPIKPLHAHAKCDSPFGLOHYTDKIIMK 119
/ Db 200 LGL-----SYPSKEIKIDOLAHQKDMYHF--PRAMTHEDVDYDPSFGLSKSAFINLVHN 250
/
/ Qy 120 -KEKEGIEKQGLISSADVIATVQHTMACHVKRTHRAILFCRQRDLLPONNAV--LVA 176
/ Db 251 AQQKREEDLDKN-----DLAASFASVYDVLTITLRA--C-----QNYPKOLVV 293
/
/ Qy 177 SGVAVSNFYIRRALEILTNA--TQCTLLCPPEPLCTDNGIMI 216
/ Db 294 AGVAVANQGLREGQLAALSAKLPVEVLVIFPLRLGDNAAMI 335
/
/ RESULT 14
/ US-08-987-121A-4
/ Sequence 4, Application US/08987121A
/ GENERAL INFORMATION:
/ APPLICANT: Hoskins, Jo Ann
/ APPLICANT: Tang, Joseph Chou-Chung
/ APPLICANT: Treadway, Patti Jean
/ TITLE OF INVENTION: Streptococcus Pneumoniae Gene Sequence
/ TITLE OF INVENTION: GCP
/ NUMBER OF SEQUENCES: 5
/ CORRESPONDENCE ADDRESS:
/ ADDRESSER: Eli Lilly and Company
/ STREET: Lilly Corporate Center
/ CITY: Indianapolis
/ STATE: Indiana
/ COUNTRY: U.S.
/ ZIP: 46285
/
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/987,121A
/ FILING DATE:
/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Webster, Thomas D.
/ REGISTRATION NUMBER: 39,872
/ REFERENCE/DOCKET NUMBER: X-
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 317-276-3334
/ INFORMATION FOR SEQ ID NO: 4:
/ SEQUENCE CHARACTERISTICS:
/

```

```

: LENGTH: 336 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-987-121A-4

Query Match      18.3%, Score 254; DB 3; Length 336;
Best Local Similarity 32.4%; Pred. No. 4,6e-22;
Matches 73; Conservative 36; Mismatches 86; Indels 30; Gaps 8;

OY 1 MEAHLTRLTNNKVEPPFLVLLIGGHCLLALVQGVSDPFLIGKSLDIAPDMIDKARR 60
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 115 MAGHMAAQSVEPLEFPLIALVLVSGHTLTVYSEADGKYKIVGETRDAVEADVGRV 174
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 61 LSLIKHEPCSTWMSGKAIENHAKGNRPHPIKRPPLHAKKCDSPFGIQTVDKILMK 120
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 175 MGL-----TYPAGREIDELAHQGHDI-YDFPAMKEDULESPSGLSKAFINLHNA 226
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 121 EKEGIEKGQILSSAADIATVQGTWMACHLVKRTFRALLPKORDLLEPQNNAVLVASGV 180
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 227 E-----QKGEISLT-EDLCASFQAAVMDIMAKTKKL-----EKYVK--TLVAVAGV 272
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 181 ASNFYIRRALEILTNATQCTLLCPPRLCTDNGIMIA-----WN 219
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 273 AANKGLRRL--ATEITDVNIIPRLRCGNAGMIAYASVSEWN 315
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 15
US-08-961-083-52
: Sequence 52, Application US/08961083
: Patent No. 6159469
: GENERAL INFORMATION:
: APPLICANT: Choi et. al.
: TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines
: NUMBER OF SEQUENCES: 452
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Human Genome Sciences, Inc.
: STREET: 9410 Key West Avenue
: CITY: Rockville
: STATE: Maryland
: COUNTRY: USA
: ZIP: 20850
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
: COMPUTER: HP Vectra 486/33
: OPERATING SYSTEM: MSDOS version 6.2
: SOFTWARE: ASCII Text
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/961,083
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Brookes, A. Anders
: REGISTRATION NUMBER: 36,373
: REFERENCE/DOCKET NUMBER: PB340P2
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (301) 309-8504
: TELEFAX: (301) 309-8512
: INFORMATION FOR SEQ ID NO: 52:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 335 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-961-083-52

Query Match      18.1%; Score 250; DB 3; Length 335;
Best Local Similarity 32.0%; Pred. No. 1.4e-21;
Matches 72; Conservative 36; Mismatches 87; Indels 30; Gaps 8

```

```

Qy 1 MEAHALTITRLTNKVEFPFLVLLISGGHCLLALVGVSDPFLIGKSLDIAFGDMLDKVAR 60
Db 114 MAGHIMAAQSYVEPLFPLALLVSGHTELIVYSEAGDYKIIVGETRDDAVGEAYDKVGRV 173
Qy 61 LSLIKHPECSTMSGKATIEHLAKOGRPHFDIKPPLHAKNCDSPFGIQTDTKIMCK 120
Db 174 MGL-----TYPAGREIDELAHQODI-YDPPRAMIKEDNLEFSFSGLKSAPINLHNA 225
Qy 121 EKEGIEKQIILSSNADIATVQHTMACHLVKTRAILFCQORDLLPONNAVLVASGV 180
Db 226 E-----QKGSLSLST-EDLCASFQAAMVDIIMAKTKKAL-----EKYPVK--ILVVAGSV 271
Qy 181 ASNFYIRRALEILTNAQCTLLCPPRLCTDNGIMIA-----WN 219
Db 272 AANKGLRERL--AAEITDVKYIIPPLRLCGDNAGMIAYASVSXWN 314

```

Search completed: November 10, 2005, 16:54:34
 Job time : 21.4761 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: November 10, 2005, 16:48:13 ; Search time 601.911 Seconds
(without alignments)
(2625.922 Million cell updates/sec)

Title: US-10-649-273-2_COPY_148_414
Perfect score: 1385
Sequence: 1 MEAHALITRLTNKVFPPLV.....DISKEVGASIKVQLKMEI 267

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 4390206 seqs, 2959870667 residues
Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgnr_1/USPRO.spool/p/US10649273/runat_02112005_091337_15540/app.query.fasta.1.1429
-DB=N_Geneseq_16Dec04 -QFMT=fastcap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bites -START=1 -END=1 -MATRIX=bloum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pro -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USBR=US10649273 @CGN 1.1 1063 @runat_02112005_091337_15540 -NCPU=6 -ICPU=3
-NO_MMAR -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOC
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: N_Geneseq_16Dec04:*
2: geneseqn1980s:*
3: geneseqn1990s:*
4: geneseqn2000s:*
5: geneseqn2001as:*
6: geneseqn2001bs:*
7: geneseqn2002as:*
8: geneseqn2002bs:*
9: geneseqn2003as:*
10: geneseqn2003bs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	1385	100.0	1416	8	ABX70950	ABX70950	Novel hum
2	1385	100.0	1526	6	ABX76635	ABX76635	DNA encod
3	1385	100.0	2058	6	ABA93268	ABA93268	Human O-s
4	1385	100.0	2197	6	ABS76635	ABS76635	DNA encod
5	1385	100.0	2572	8	ABT23207	ABT23207	Human pro

6	1348	97.3	1820	6	AA046856	AA046856	Human gly
7	1348	97.3	1820	8	ACA60887	ACA60887	Human cDN
8	1348	97.3	1821	10	AB557020	AB557020	CDNA encd
9	1204	86.9	2208	10	ADA52832	ADA52832	Human cod
10	1204	86.9	2890	12	ADQ24627	ADQ24627	Human sof
11	995.5	71.9	3358	10	AD531345	AD531345	Human dia
12	983	71.0	1572	6	AB075508	AB075508	Human si
13	599	43.2	2734	5	AA584622	AA584622	DNA encod
14	468	33.8	371	12	ADL86725	ADL86725	DNA up-re
15	468	33.8	371	12	ADL86726	ADL86726	DNA up-re
16	415.5	30.0	1601	4	AB124633	AB124633	Drosophi1
17	415.5	30.0	3656	3	AB124632	AB124632	Drosophi1
18	401.5	29.0	1557	3	AA038454	AA038454	Arabidops
19	342	24.7	4360	6	AA048239	AA048239	Ehrlichia
20	338.5	24.4	1385	6	AAH15110	AAH15110	Human cDN
21	321.5	23.2	1146	8	ACA26804	ACA26804	Prokaryot
22	314.5	22.7	936	8	ACA20445	ACA20445	Prokaryot
23	308.5	22.3	1044	8	ACA39102	ACA39102	Prokaryot
24	308.5	22.3	94750	4	AAE28551	AAE28551	Genomic f
25	307	22.2	1053	12	ADL03120	ADL03120	DNA encod
26	304.5	22.0	1032	8	ACA43173	ACA43173	Prokaryot
27	301.5	21.8	1000	4	AAE91424	AAE91424	Moraxella
28	301.5	21.8	1000	6	ABK37804	ABK37804	DNA seque
29	300	21.7	1092	6	ABQ90383	ABQ90383	M. capsul
30	291.5	21.0	1026	8	AA54064	AA54064	Pseudomon
31	291.5	21.0	1026	8	ACA42146	ACA42146	Pseudomon
32	291.5	21.0	1026	10	ADG73341	ADG73341	P aerugin
33	291.5	21.0	1059	11	ABD02280	ABD02280	Pseudomon
34	291.5	21.0	1206	11	ABD02197	ABD02197	P aerugin
35	289.5	20.9	1026	10	ADG73343	ADG73343	P aerugin
36	289.5	20.9	1029	4	AA553309	AA553309	Haemophi1
37	289.5	20.9	1029	8	ACA34150	ACA34150	Haemophi1
38	285	20.6	110000	2	AA742063	AA742063	Continuati
39	285	20.6	372	5	ABV61069	ABV61069	Human pro
40	282.5	20.4	9667	13	ADT05493	ADT05493	Haemophi1
41	282.5	20.4	85814	13	ADT05464	ADT05464	Haemophi1
42	280.5	20.3	1020	8	ACA44384	ACA44384	Prokaryot
43	280.5	20.3	1074	10	ADP02056	ADP02056	Bacteri
44	279.5	20.2	1014	4	AA556045	AA556045	Salmone11
45	279.5	20.2	1014	8	ACA51431	ACA51431	Prokaryot

ALIGNMENTS

RESULT 1	ABX70950	standard; cDNA, 1416 BP.
ID	ABX70950	
AC	ABX70950	
DT	05-MAR-2003	(first entry)
DE	Novel human cDNA sequence #175.	
XX	Human; gene; ss; nervous system disorder; peripheral neuropathy;	
XX	Huntington's disease; amyotrophic lateral sclerosis; hemophilia;	
XX	neurodegenerative disease; Parkinson's disease; Alzheimer's disease;	
XX	autoimmune disease; systemic lupus erythematosus; rheumatoid arthritis;	
XX	insulin-dependent diabetes mellitus; anaemia; thrombocytopenia; wound;	
XX	ulcer; burn; bone disorder; osteoporosis; osteoarthritis; stroke;	
XX	fibrosis; reperfusion injury; infection; allergic rhinitis; asthma;	
XX	coagulation disorder; cancer; tumour; inflammatory disease; septic shock;	
XX	Crohn's disease; anapylaxis; proliferation; chemotactic;	
XX	differentiation; stem cell growth factor; haematopoiesis; chemokinetic;	
XX	haemostatic; antiinflammatory; expressed sequence tag; EST.	
OS	Homo sapiens.	
XX	PN	WO200281731-A2.
XX	PD	17-OCT-2002.
XX	PF	29-JUN-2002; 2002WO-US001222.

XX 30-JAN-2001; 2001US-00774528.
 PR (HXS-) HXS INC.
 PA (GOOD/) GOODRICH R W.
 PA
 XX Tang TY, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
 PI Xie AJ, Yang Y, Wehrman T, Wang J, Wang D, Drmanac RT;
 XX WPI: 2003-058563/05.
 DR
 XX Novel polypeptide useful for treating neurodegenerative diseases, myeloid
 PT or lymphoid cell disorders, bone disorders, mechanical and traumatic
 PT disorders, coagulation disorders, and inflammatory diseases.
 XX
 PS Claim 1; Page: 612pp; English.

CC This invention relates to the cDNA sequences encoding an isolated novel
 CC human polypeptide. The protein encoded by the nucleic acid of the
 CC invention is useful for treating central and peripheral nervous system
 CC diseases (e.g. peripheral neuropathy, Huntington's disease, amyotrophic
 CC lateral sclerosis); neurodegenerative diseases (e.g. Parkinson's disease,
 CC Alzheimer's disease); autoimmune disease (e.g. systemic lupus
 CC erythematosus, rheumatoid arthritis, insulin-dependent diabetes mellitus)
 CC ; myeloid or lymphoid cell disorders (e.g. anaemia and thrombocytopenia)
 CC ; wounds, ulcers, burns; bone disorders (e.g. osteoporosis,
 CC osteoarthritis); mechanical and traumatic disorders (e.g. stroke, head
 CC trauma); lung or liver fibrosis; reperfusion injury in various tissues;
 CC bacterial, viral or fungal infections; allergic conditions such as
 CC allergic rhinitis, asthma; coagulation disorders (e.g. haemophilia);
 CC cancer and tumours; and inflammatory diseases (e.g. septic shock, Crohn's
 CC disease, anaphylaxis). The protein may be used to inhibit the growth,
 CC infection or function of infectious agents such as bacteria, fungi,
 CC viruses, or to effect bodily characteristics, biorythms or circadian
 CC cycles of rhythms. The protein may also have
 CC proliferation/differentiation, stem cell growth factor, haematopoiesis
 CC regulation, immune stimulation or suppressing, chemotactic/chemokinetic,
 CC haemostatic and thrombolytic, receptor/ligand, and antiinflammatory
 CC activities. The cDNA sequences of the invention are useful for expressing
 CC recombinant protein for analysis. The present sequence represents a novel
 CC human cDNA sequence of the invention, this sequence is an expressed
 CC sequence tag (EST) and was identified using subtractive hybridisation
 XX
 SO Sequence 1416 BP; 441 A; 280 C; 273 G; 422 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 4.68e-147 Length: 1416
 Score: 1385.00 Matches: 267
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 8 Gaps: 0

US-10-649-273-2_COPY_148_414 (1-267) X ABX70950 (1-1416)

QY 1 MetGUAUAHSAIAleuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20
 DB 502 ATGGAGGCTCATGCACTTAATTAGGTGACCAATTAAGATTTCTTTTAAAGT 561
 QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlyValSerAspPheLeu 40
 DB 562 CTTTGTGATTTCTGGAGGCTCAGTCTGTTGGCATTAAGTTCAAGAGATTTCTG 621
 QY 41 LeuLeuGlyIleSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60
 DB 622 CTTCTTGAAGAGCTTTTGGACATAGACCAAGTGACATGCTTGCACAAAGGGAAGA 681
 QY 61 LeuSerLeuIleLysHisProGlyCysSerThrMetSerGlyGlyLysAlaIleGluHis 80
 DB 682 CTTTCTTAATTAACAATCAGAGTGCTCCACCATAGAGTGAGGAAAGCAATAGAAAT 741
 QY 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisAlaLys 100

DB 742 TTGGCCAAACAGAGAAATGATTTGATTTGACATCAAACTCCCTGCATCATGCTAAA 801
 QY 101 AenCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120
 DB 802 AATTGTGATTTTCTTTTACTGAGCTTCAACGTTACTGATTAATAATTAATGAAAAAG 861
 QY 121 GlnLysGlnGlnGlyIleGlnLysGlyGlnIleLeuSerSerAlaAlaAspIleAla 140
 DB 862 GAAAAAGAGAGATTTAGAGAGGGCAAACTCTGCTTTCAGACAGACATTCGCTGCC 921
 QY 141 ThrValGlnHisIleThrMetAlaCysHisIleuValLysArgThrHisArgAlaIleLeuPhe 160
 DB 922 AAGATACAGACACACATGCGCATGCTTGGAAAAAGAACATCGGCGTATTCGT 961
 QY 161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyLysVal 180
 DB 982 TGTAGCAGAGAGACTGTTACTCTCAAAATTAATGAGTACTGTTGCATCTGGTGGTGC 1041
 QY 181 AlaSerAsnPheThrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThr 200
 DB 1042 GGAAGTAATCTTATATATCCGAGAGCTTGAAATTTTAACAAACGCAACACAGTGCAC 1101
 QY 201 LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaTPanGly 220
 DB 1102 TTGTTGTCTCTCTCCCTCCAGACTATGCACTGATTAATGGCATTAATGATTCGATGAAATGGT 1161
 QY 221 IleGlnArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlnGlyIleArgTyrGlu 240
 DB 1162 ATTTGAAAGACTGCTGCTGCTGGCATTTTATACATGACATGAGAGCATCCCTATGAA 1221
 QY 241 ProLysCysProLeuGlyValAspIleSerLysGlyValGlyGlnAlaSerIleLysVal 260
 DB 1222 CCAAAATGCTCTTGTGAGTATGACATATCAAAAGAGTTGGAAGACTTCATTAAGTA 1281
 DB 261 ProGlnLeuLysMetGluIle 267
 DB 1282 CCACATTAATAAATGAGAGATA 1302
 QY
 DB
 AC
 XX
 DT 11-DEC-2002 (first entry)
 DE
 XX
 DE DNA encoding novel human metalloprotease MPI fragment #1.
 XX
 KW Metalloprotease; MP-1; immune disorder; glutamate transport; cancer;
 KW motor neuron disorder; amyotrophic lateral sclerosis; ALS; diabetes;
 KW reproductive disorder; Klinefelter's syndrome; germinal cell aplasia;
 KW genital wart; metabolic disorder; premature puberty; Kallman syndrome;
 KW Cushing's syndrome; neurodegenerative disease; Alzheimer's disease;
 KW Parkinson's disease; Huntington's disease; Tourette syndrome; sepsis;
 KW liver disease; renal disease; immune disorder; rheumatoid arthritis;
 KW acquired immunodeficiency syndrome; AIDS; pulmonary disease; pneumonia;
 KW emphysema; cystic fibrosis; vascular disorder; inflammatory disorder;
 KW neurological disorder; gene; ds.
 KW
 XX
 OS Homo sapiens.
 XX
 PN MO200272751-A2.
 XX
 PD 19-SEP-2002.
 XX
 PF 05-FEB-2002; 2002MO-US003353.
 XX
 PR 05-FEB-2001; 2001US-0266518P.
 PR 10-APR-2001; 2001US-0282814P.
 XX
 PA (BRIM) BRISTOL-MYERS SQUIBB CO.
 XX
 PI Chen J, Feder J, Nelson TC, Duclos F, Krystek S;

XX MPI, 2002-723329/78.
DR F-PSDB; ABG96487.
XX
XX New isolated nucleic acid encoding MP-1 protein, useful for preventing,
PT treating, or ameliorating diseases associated with aberrant
PT metalloproteinase activity, e.g. immune, metabolic, inflammatory and
PT neurological disorders.
XX
XX Disclosure; Page 462-463; 473pp; English.
XX
XX The invention describes an isolated nucleic acid molecule (1) encoding a
CC metalloproteinase (MP-1). (1) is useful for preventing, treating, or
CC ameliorating a medical condition, particularly an immune disorder, an
CC aberrant glutamate transport or motor neuron disorder, such as
CC amyotrophic lateral sclerosis (ALS), its juvenile form or an ALS-like
CC condition. The compositions and methods are also useful for diagnosing,
CC prognosticating, treating, ameliorating and/or treating disorders
CC associated with MP-1 activity, e.g. diabetes, cancer, reproductive
CC disorders (e.g. Klinefelter's syndrome, genital warts, or germinal cell
CC aplasia), metabolic disorders (e.g. premature puberty, Kallman syndrome,
CC or Cushing's syndrome), neurodegenerative diseases (Alzheimer's disease,
CC Parkinson's disease, Huntington's disease or Tourette syndrome), liver
CC and renal diseases and immune disorders (e.g. AIDS, rheumatoid arthritis
CC or sepsis), pulmonary diseases (e.g. pneumonia, emphysema or cystic
CC fibrosis) and vascular, inflammatory and neurological disorders (e.g.
CC Alzheimer's disease or Parkinson's disease). This sequence represents a
CC metalloproteinase MP1 polynucleotide
XX
SQ Sequence 1526 BP; 484 A; 297 C; 289 G; 456 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 5.2e-147 Length: 1526
Score: 1385.00 Matches: 267
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-649-273-2_COPY_148_414 (1-267) x ABS76639 (1-1526)
QY 1 MetGluAlaHisAlaLeuThr11eArgLeuThrAsnLysVal1GluPheProPheLeuVal 20
DB 1 ATGGAGGCTCATGCACTTACTATTAGGTTGCCAATAAGTAGAATTTCTTTTATGTT 60
QY 21 LeuLeu11eSerGlyGlyHisCysLeuLeuAlaLeuVal1GlnGlyValSerAspPheLeu 40
DB 61 CTTTGAATTTCTGAGGTCACTGCTGTGTGGCATTTAGTTCAGGAGTTTCGATTTCTTG 120
QY 41 LeuLeuG1LysSerLeuAsp11eAlaProG1LysPheLeuAspLysVal1AlaArgArg 60
DB 121 CTTCTTGGAAGCTTTTGGACATAGCACAGTACATCTCTTGACAGAGTGGCAAGAGA 180
QY 61 LeuSerLeu11eLysHisProG1LysSerThrMetSerGlyGlyLysAla11eGluHis 80
DB 181 CTTTCTTTAATAAACAATCCAGAGTGTCTCCACCATGATGATGGTGAAGCATGAACAT 240
QY 81 LeuAlaLysGlnGlyAsnArgPheHisPheAsp11eLysProPoleuHisHisAlaLys 100
DB 241 TTGGCCAAACACAGGAATATGATTTTCATTTGACATCAACTCCCTTGCATCATGCTAAA 300
QY 101 AsnCysAspPheSerPheThrG1LysGlnHisVal1ThrAspLys11eMetLysLys 120
DB 301 AATTGATTTTCTTTTACTGACCTTCAACAGCTTACTGTAATAATATATATGAAAAG 360
QY 121 G1LysG1LysG1Lys11eG1LysG1Lys11eLeuSerSerAla11eAlaAsp11eAla 140
DB 361 GAAAAGAGAGAGATNTAGAGAGGGCCAAATCTCTTTCAGACAGACATTTGCTGCC 420
QY 141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAla11eLeuPhe 160
DB 421 ACAGTACAGACACAAATGCGATGTCATCTGTGTAAAGAACATCGGGCTATTCTGTTT 480

QY 161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaVal1LeuVal1AlaSerG1LysVal 180
DB 481 TGTAAAGACAGAGACTTGTACTCTAAATAATGACGACTGCTGTCATCTGTGTGTC 540
QY 181 AlaSerAsnPheThr11eArgArgAlaLeuGlu11eLeuThrAsnAlaThrGlnCysThr 200
DB 541 GCAAGTACTTTATATCCGACAGCTCTGGAAATTTTACAAACCAACACACAGTCACT 600
QY 201 LeuLeuCysProProProAlaArgLeuCysThrAspAsnG1LysMet11eAlaTPaenGly 220
DB 601 TTGTTGTGTCCTTCCCTCCAGACTATGATCATGATATATGGCATTTATGATTCATGGAATGG 660
QY 221 11eGluArgLeuArgAlaLysLysGly11eLeuHisAsp11eG1LysArg11eArg11e 240
DB 661 ATTTGAAGACTACGCTGCTGGCTTGGCATTTTACATGACATGAAAGCATCCGCTATGAA 720
QY 241 ProLysCysProLeuG1LysAlaPheSerLysGluVal1G1LysAlaSer11eLysVal 260
DB 721 CCAAAATGCTCTTGTGAGTATGACATATCAAAAGAAATTGAGAGAGCTTCATAAAGTA 780
QY 261 ProGlnLeuLysMetGlu11e 267
DB 781 CCACATTTAAATAATGAGATA 801

RESULT 3
ABA93268
ID ABA93268 standard; cDNA; 2058 BP.
XX
XX ABA93268;
XX
XX 19-APR-2002 (first entry)
XX
XX
XX Human O-sialoglycoproteinase-like protein encoding cDNA SEQ ID NO:1.
DE
XX
XX Human; O-sialoglycoproteinase-like protein; OSGPLP; enzyme; gene; ss.
OS
XX Homo sapiens.
XX
XX
XX Key Location/Qualifiers
FT CDS 110..1354
FT /tag= a
FT /product= "O-sialoglycoproteinase-like protein"
XX
XX CNA138550-A.
XX
XX 24-OCT-2001.
XX
XX 19-APR-2000; 2000CN-00106834.
PF
XX 19-APR-2000; 2000CN-00106834.
PR
XX (SHAN-) SHANGHAI SHENGYUAN GENE DEV CO LTD.
PA
XX
XX Mao Y, Xie Y;
PI
XX
XX MPI, 2002-115090/16.
DR
XX P-PSDB; ABB05481.
DR
XX
XX O-sialoglycoproteinase-like protein and encoding polynucleotide, useful
PT for diagnosing, preventing and treating related diseases.
PT
XX
XX
XX Claim 5; Page 29-30 (Disclosure); 38pp; Chinese.
PS
XX
XX The present sequence encodes human O-sialoglycoproteinase-like protein
CC (OSGPLP). The present invention also describes: (1) the preparation of
CC the OSGPLP protein; (2) applying the OSGPLP protein in diagnosis; (3) the
CC prevention and/or treatment of related diseases; (4) utilizing the OSGPLP
CC protein in screening its agonist, excitomotor and inhibitor and preparing
CC an antibody against the OSGPLP protein; and (5) the use of the OSGPLP
CC polynucleotide sequences, proteins, agonists, excitomotors, inhibitors
CC and antibodies in treating diseases related to the abnormal OSGPLP gene
CC and in preparing the medicine composite for the treatment
XX

SQ Sequence 2058 BP; 637 A; 400 C; 410 G; 611 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	7,97e-147	Length:	2058
Score:	1385.00	Matches:	267
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-10-649-273-2_COPY_148_414 (1-267) x ABS76635 (1-2058)

```
QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20
Db 551 ATGGAGGCTCATGCACTTACTATTAGTTGACCATTAAGTATGATTTCTTTTATGTT 610
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
Db 611 CTTTGGATTTCGTGGAGGCTGCTGCTGTGGCATTTAGTCAAGGAGTTTCAGATTTCG 670
QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60
Db 671 CTTCTTGGAAAGCTTTGGACATAGCACAGGTGACATGCTTGACAGGTGGCAAGAA 730
QY 61 LeuSerLeuIleLysHisProGlyCysSerThrMetSerGlyGlyValAlaIleGluHis 80
Db 731 CTTTCTTTAATTAACAATCCAGAGTCTCCACCATGATGGTGGGAAAGCCATAGACAT 790
QY 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisAlaLys 100
Db 791 TTGGCCAAACAAAGAAATAGATTTCATTTTGGACATCAACCTCCCTGCATCATGCTAAA 850
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValIleThrAspLysIleIleMetLysLys 120
Db 851 AATTGTGATTTTCTTTACTGAGCTTCAACACGTTACTGATTAATAATATGAAAAAG 910
QY 121 GluLysGlnGluGlyIleGlyLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaAla 140
Db 911 GAAAGAGGAAAGGATTGAGAAAGGGGCAATCTGCTTTCAGCAGCAGCATTTGCTGCC 970
QY 141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160
Db 971 ACACTACAGCACACATGCAATGTCATCTTGGAAAAAGAACATCGGGCTATTCGTTT 1030
QY 161 CysLysGlnIleArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyVal 180
Db 1031 TGTAAAGCAGAGAGACTTGTACTCTCAAAATATATGACATGCTGTGCATCTGTGTGTC 1090
QY 181 AlaSerAsnPheTyrlleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThr 200
Db 1091 GCAAGTAACTTCTATATCCGAGAGCTCTGGAAATTTTAAACAAAGCCACACAGTGCCT 1150
QY 201 LeuLeuCysProProCysProArgLeuCysThrAspAsnGlyIleMetIleAlaTrpAsnGly 220
Db 1151 TTGTGTGTCTCTCCCAAGACTATGCACTGATTAATGGATATATGATTCATGAGATGCT 1210
QY 221 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlnGlyIleArgTyrlGlu 240
Db 1211 ATTGAAGAAGCTACGTCGTGGCTTGGCATTTTAACTGACATGAAGAGCTCCGCTATGAA 1270
QY 241 ProLysCysProLeuGlyValAspIleSerLysGlyValGlyAlaSerIleLysVal 260
Db 1271 CCAAAATGTCTCTTGGAGTGAACATATCAAAAGAAAGTTGGAAAGCTTCCATTAAGTA 1330
QY 261 ProGlnLeuLysMetGluIle 267
Db 1331 CCAACATTAATAATGACATA 1351
```

RESULT 4
ABS76635
ID ABS76635 standard; DNA; 2197 BP.

XX ABS76635;

XX 11-DEC-2002 (first entry)

DE DNA encoding novel human metalloprotease MP1.

KW Metalloprotease; MP-1; immune disorder; glutamate transport; cancer;
KW motor neuron disorder; amyotrophic lateral sclerosis; ALS; diabetes;
KW reproductive disorder; Klinefelter's syndrome; germinal cell aplasia;
KW genital wart; metabolic disorder; premature puberty; Kallman syndrome;
KW Cushing's syndrome; neurodegenerative disease; Alzheimer's disease;
KW Parkinson's disease; Huntington's disease; Tourette syndrome; sepsis;
KW liver disease; renal disease; immune disorder; rheumatoid arthritis;
KW acquired immunodeficiency syndrome; AIDS; pulmonary disease; pneumonia;
KW emphysema; cystic fibrosis; inflammatory disorder; inflammatory disorder;
KW neurological disorder; gene; ds.

OS Homo sapiens.

PN WO200272751-A2.

XX 19-SEP-2002.

PF 05-FEB-2002; 2002MO-US003353.

XX 05-FEB-2001; 2001US-0266518P.

PR 10-APR-2001; 2001US-0282814P.

XX (BRIM) BRISTOL-MYERS SQUIBB CO.

PI Chen J, Feder J, Nelson TC, Duclos F, Krystek S;

XX MPI; 2002-723329/78.

DR P-Psdb; ABS66478.

PT New isolated nucleic acid encoding MP-1 protein, useful for preventing,
PT treating, or ameliorating diseases associated with aberrant
PT metalloprotease activity, e.g. immune, metabolic, inflammatory and
PT neurological disorders.

XX Claim 1; Fig 1A-C; 473pp; English.

XX The invention describes an isolated nucleic acid molecule (I) encoding a
CC metalloprotease (MP-1). (I) is useful for preventing, treating, or
CC ameliorating a medical condition, particularly an immune disorder, an
CC aberrant glutamate transport or motor neuron disorder, such as
CC amyotrophic lateral sclerosis (ALS), its juvenile form or an ALS-like
CC condition. The compositions and methods are also useful for diagnosing,
CC prognosticating, treating, ameliorating and/or treating disorders
CC associated with MP-1 activity, e.g. diabetes, cancer, reproductive
CC disorders (e.g. Klinefelter's syndrome, genital warts, or germinal cell
CC aplasia), metabolic disorders (e.g. premature puberty, Kallman syndrome,
CC or Cushing's syndrome), neurodegenerative diseases (Alzheimer's disease,
CC Parkinson's disease, Huntington's disease or Tourette syndrome), liver
CC and renal diseases and immune disorders (e.g. AIDS, rheumatoid arthritis
CC or sepsis), pulmonary diseases (e.g. pneumonia, emphysema or cystic
CC fibrosis) and vascular, inflammatory and neurological disorders (e.g.
CC Alzheimer's disease or Parkinson's disease). This sequence represents a
XX metalloprotease MP1 polynucleotide

SQ Sequence 2197 BP; 681 A; 441 C; 439 G; 636 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	8,75e-147	Length:	2197
Score:	1385.00	Matches:	267
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-10-649-273-2_COPY_148_414 (1-267) x ABS76635 (1-2197)

```
QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20
|||
```

Db 672 ATGAGGCTCATGCACTTACTATTAGCTTGACCAATAAGTAGAATTCCTTTTACTT 731
QY 21 LeuLeuIleSerGlyGlyYHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
Db 732 CTTTGTGATTTCTGGAGGCTCATGCTCTGTGGCATTAGTTCAGAGGATTTTTCAGATTTTCTG 791
QY 41 LeuLeuGlyIleSerSerLeuAspIleAlaProGlyAspMetLeuAspIleValAlaArgArg 60
Db 792 CTTCTTGAAAGCTCTTGGACATAGCACAGGTGCATGCTTGACAAAGGTGCAGAAAGA 851
QY 61 LeuSerLeuIleYHisIleProGlyCysSerThrMetSerGlyGlyValAlaIleGlnHis 80
Db 852 CTTTCTTTAATAAATCATCCAGAGTCTCCACATAGAGTGGTGGAAACCCATAGAACAT 911
QY 81 LeuAlaIleGlnGlyAsnArgPheHisIleAspAspIleYSerProProLeuHisIleAlaYAs 100
Db 912 TTGGCCAAACAAAGAAATAGATTTTTCATTTTGACATCAACCTCCCTTGATCATGCTAA 971
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspIleIleIleMetIleYAs 120
Db 972 AATTGTATTTTCTTTTACTGCACTTCACACGTTACTGATTAATAATATGAAAAAG 1031
QY 121 GlnIleGlyGlnGlyIleGlnIleGlyGlnIleLeuSerSerAlaAlaAspIleAlaAla 140
Db 1032 GAAAAAGGAGAGGATTTAGAGAGGGGCAAAATCCGTCTTACGACGACATTCCTGCT 1091
QY 141 ThrValGlnHisThrMetAlaCysHisIleValIleValIleValIleValIleValIle 160
Db 1092 ACGTACACGACACACATGCGATGCTCTTGTGAAAAAACAATCGGGCTATTTCTGTTT 1151
QY 161 CysIleGlnArgAspLeuLeuProGlnAsnAsnAlaValIleValAlaSerGlyGlyVal 180
Db 1152 TGTAAAGAGAGAGACTTGTACTCAAAATATATGACATGCTGTGATCTGTGCTGTC 1211
QY 181 AlaSerAsnPheIleIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThr 200
Db 1212 GCAAGTAACTTCTATATCCGAGAGCTCTGAAAAATTTTAAACAAACAGACAGTCACT 1271
QY 201 LeuLeuCysProProArgLeuCysThrAspAsnGlyIleMetIleAlaIleArgAsnGly 220
Db 1272 TTGTGTGTCTCTCCACGACTTACCTATATATGAGCATTTATATGATGATGATGATGAT 1331
QY 221 IleGlnArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlnGlyIleArgTyrGln 240
Db 1332 ATTGAAGACTACCTGCTGGCATTTTACATGACATAGAGAGATCCGCTATGTA 1391
QY 241 ProIleCysProLeuGlyValAspIleSerIleGlyValIleGlyValIleSerIleYVal 260
Db 1392 CCAAAATGCTCTTGGAGTACATATCAAAAGAAATTGAGAGAGCTTCCATAAAGTA 1451
QY 261 ProGlnLeuYMetGlnIle 267
Db 1452 CCACAATTAATAAATGAGATA 1472

RESULT 5
ABT23207
ID ABT23207 standard; DNA; 2572 BP.
AC ABT23207;
XX
XX 01-MAY-2003 (first entry)
XX
XX Human protein modification + maintenance molecule DNA SEQ ID No 36.
XX
XX Cytosolic; antiarteriosclerotic; anti-HIV; antiallergic; anticonvulsant;
XX cerebroprotective; antiparkinsonian; nootropic; antiinflammatory;
XX antitumor; hepatotropic; gynaecological; antibacterial; vincidine;
XX protozoacide; antiparasitic; cell proliferative disease; PMOD;
XX protein modification and maintenance molecule; immunogenic fragment;
XX cancer; autoimmunity; inflammatory disease; neurological disorder;
XX gastrointestinal; developmental; vesicle trafficking disorder; infection;
XX protein-protein interaction; drug-target interaction;
XX gene expression profile; human; gene; ds.

XX OS Homo sapiens.
XX XX
XX W0200300844-A2.
XX
XX
XX 03-JAN-2003.
XX
XX
XX 18-JUN-2002; 2002MO-US019360.
XX
XX 22-JUN-2001; 2001US-0300508P.
XX 06-JUL-2001; 2001US-0303445P.
XX 13-JUL-2001; 2001US-0305405P.
XX 09-AUG-2001; 2001US-0311442P.
XX 24-AUG-2001; 2001US-0314821P.
XX 29-AUG-2001; 2001US-0315992P.
XX 03-MAY-2002; 2002US-0378205P.
XX
XX
XX (INCY-) INCYTE GENOMICS INC.
XX
XX Gandhi AR, Kable AE, Swarnakar A, Hafalia AJA, Tran B, Duggan BM,
XX Warren BA, Ison CH, Honchell CD, Nguyen DB, Lu DAM, Lee EA, Yue H;
XX Forrester JF, Barroso I, Ramkumar J, Griffin JA, Li JX, Yang J;
XX Thangavelu K, Gietzen KJ, Ding L, Baughn MR, Borowsky ML, Yao MG;
XX Walla NK, Mason PM, Gururajan R, Lee S, Becha SD, Lee ST, Tran UK,
XX Elliott VS, Luo W, Sprague MW, Tang YT, Lu Y, Zebardjian Y;
XX
XX WPI; 2003-184039/18.
XX P-PSDB; ABJ26654.
XX
XX New isolated human PMOD polypeptide and polynucleotide, useful for
XX diagnosing, treating and preventing diseases or conditions associated
XX with the aberrant PMOD expression, e.g. cancer, AIDS, atherosclerosis and
XX infections.
XX
XX Claim 91; Page 211; 225pp; English.

The invention relates to an isolated polypeptide comprising: any of 28
sequences of 48-1256 amino acids; a natural amino acid sequence at least
90% identical to the 28 amino acid sequences; 94% identical to a sequence
of 703 or 267 amino acids, 96% identical to a sequence of 414 amino
acids, or 97% identical to a sequence of 242 amino acids, all given in
the specification; or a biologically active or immunogenic fragment of
the isolated polypeptide. The polypeptides and polynucleotides are useful
in diagnosing, treating and preventing diseases or conditions associated
with the decreased expression of protein modification and maintenance
molecules (PMOD), such as cell proliferative diseases (e.g. cancer,
atherosclerosis), autoimmune/inflammatory diseases (e.g. AIDS,
allergies), neurological disorders (e.g. stroke, Parkinson's disease,
epilepsy), gastrointestinal (e.g. ulcer, cirrhosis), reproductive (e.g.
endometriosis), developmental, vesicle trafficking disorders, and
infections (e.g. bacterial, viral, parasitic, protozoal). These are also
useful in assessing the effects of exogenous compounds on the expression
of nucleic acid and amino acid sequences of PMOD. The PMOD or its
fragments are useful in screening compounds for effectiveness as agonist
or antagonist of the polypeptides, or in altering the expression of the
target polynucleotide and compounds that specifically bind to or modulate
the activity of the polypeptide. The microarray is useful in monitoring
or measuring protein-protein interactions, drug-target interactions, and
gene expression profiles. This polynucleotide sequence represents the DNA
encoding a human PMOD protein of the invention

Sequence 2572 BP; 780 A; 489 C; 509 G; 794 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 1,09e-146 Length: 2572
Score: 1385.00 Matches: 267
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 8 Gaps: 0

US-10-649-273-2_COPY_148_414 (1-267) x ABT23207 (1-2572)

QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20
| | | | |
Db 585 ATGGAGGCTCATGCACTTACTATTAGTTGACCAATAAAGATATTCCTTTTATGTT 644
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
| | | | |
Db 645 CTTTGGATTTCTGGAGGCTCACGTCCTGTTGGCATTAGTTCAAGGAGTTTCAGATTTTCTG 704
QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60
| | | | |
Db 705 CTTTGGAAAGCTCTTGGACATAGCACAGGTGACATGCTTGACAAAGGTGCAAGAAAGA 764
QY 61 LeuSerLeuIleLysHisProGluCysSerThrMetSerGlyGlyLysAlaIleGluHis 80
| | | | |
Db 765 CTTTCTTTAATTAATTAATCCAGAGCTCCACCATGATGGTGGGAAAGCCATAGAACAT 824
QY 81 LeuValLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 100
| | | | |
Db 825 TTGGCCAAACAAAGAAATAGATTTCATTTTGACATCAAACTCCCTTGATCATCTAA 884
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120
| | | | |
Db 885 AATTGTGATTTTCTTTTACTGAGCTTCAACACGTTACTGATTAATAATATGAAAG 944
QY 121 GluLysGluGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaAla 140
| | | | |
Db 945 GAAAGAAAGAAAGATTTGAGAAAGGGAATCCTGCTTCAGCAGCAGACATTCCTGCC 1004
QY 141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160
| | | | |
Db 1005 ACAGTACAGCACAACATGCAATGTCATCTTGGAAAGAACACATCGGCTATTCTGTTT 1064
QY 161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyVal 180
| | | | |
Db 1065 TGTATAGCAGACAGACTTGTACTTCAAAATATGACATGCTGTCATCTGTGTGTC 1124
QY 181 AlaSerAsnPheThrIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThr 200
| | | | |
Db 1125 GCAAGTATCTTATATCCGAGAGCTCTGGAAATTTTAAACAAAGCAACACAGTGCCT 1184
QY 201 LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaTrpAsnGly 220
| | | | |
Db 1185 TTGTGTCTCCTCCCTCCAGACTATGCACTGATTAATGATTAATGATTCATGAGATGCT 1244
QY 221 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrGlu 240
| | | | |
Db 1245 ATTGAAGAAGACTACGCTGCTGGCTTTTACATGATGATGAAGGCAATCCGCTATGAA 1304
QY 241 ProLysCysProLeuGlyValAspIleSerLysGluValGlyGluAlaSerIleLysVal 260
| | | | |
Db 1305 CCAAAATGTCTCTTGGAGTACATATCAAAAGAAAGTTGGAAAGCTTCCATAAAGTA 1364
QY 261 ProGlnLeuLysMetGluIle 267
| | | | |
Db 1365 CCACATTTAAATGACAGATA 1385

KW chromosome mapping; tissue typing; gene therapy; neuroprotective;
KW cytostatic; anorectic; cardiant; haemostatic; gene; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 146..1390
FT /tag=a
FT /product="Human 28472 protein"
FT /note="This region is specifically claimed as SEQ ID NO:
6 in claim 1 of the specification"
XX WO200274960-A2.
XX 26-SEP-2002.
XX 08-NOV-2001; 2001WO-US051427.
XX 08-NOV-2000; 2000US-0246768P.
XX 08-NOV-2000; 2000US-0246772P.
XX 15-NOV-2000; 2000US-0249185P.
XX (MILL-) MILLENNIUM PHARM INC.
XX Leiby KR, Kapeller-Libermann R, Glucksmann M;
XX WPI: 2002-759898/82.
XX P-PsDB; AAE29234.
XX New 38650, 28472, 5495, 65507, 81588 or 14354 nucleic acid molecules
XX PT useful for diagnosing and treating cancer, immune, cardiovascular,
XX PT hemotopoietic, brain, pain, metabolic, liver or platelet disorders, and
XX PT in pharmacogenomics.
XX Claim 1; Fig 8; 178pp; English.
XX PS
XX CC The present invention relates to novel 38650, 28472, 5495, 65507, 81588
XX CC or 14354 nucleic acid molecules which encode adenosine deaminase, glyco-
XX CC protease or seven transmembrane domain (7TM) receptor family members.
XX CC Sequences of the invention are useful in diagnosing and treating cancer
XX CC or aberrant cellular proliferation and/or differentiation (e.g. colon or
XX CC lung cancer), immune disorders (e.g. selective IGA deficiency, rheumatoid
XX CC arthritis, diabetes, allergy), heart or cardiovascular disorders (e.g.,
XX CC hypertension, atherosclerosis, arrhythmias, ischemic heart disease,
XX CC myocardial infarction, thrombus) including endothelial cell disorders
XX CC (e.g. psoriasis, Grave's disease), haematopoietic disorders, brain
XX CC disorders (e.g. cerebral oedema, Parkinson's or Alzheimer's disease),
XX CC pain and metabolic disorders (e.g. obesity), liver disorders or platelet
XX CC disorders. They are also useful in screening assays, predictive medicine
XX CC (e.g. diagnostic assays, prognostic assays, monitoring clinical trials
XX CC and pharmacogenetics) and prophylactic and therapeutic methods. The
XX CC nucleic acids may also be used in chromosome mapping, tissue typing and
XX CC forensic biology and as surrogate markers. Sequences of the invention are
XX CC also used in gene therapy. The present sequence is human glycoprotease
XX CC 28472 CDNA
XX SQ Sequence 1820 BP; 543 A; 365 C; 393 G; 518 T; 0 U; 1 Other;
XX
XX Alignment Scores:
XX Pred. No.: 1,07e-142 Length: 1820
XX Score: 1348.00 Matches: 260
XX Percent Similarity: 98.50% Conservative: 3
XX Best Local Similarity: 97.38% Mismatches: 4
XX Query Match: 97.33% Indels: 0
XX DB: Gaps: 0
XX
XX US-10-649-273-2_COPY_148_414 (1-267) x AAD46856 (1-1820)
QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20
| | | | |
Db 587 ATGGAGGCTCATGCACTTACTATTAGTTGACCAATAAAGATATTCCTTTTATGTT 646
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40

QY 81 LeuAlaIaIySGInGIyAaNaArgPheHisPheAspIleIySPProPLeuHisIaIaIyS 100
DB 828 TTGGCCCAAGAGAAATAGATTTCATTTGACATCAAACTCCCTTCATCATCTAA 887
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspIySIIleIleMetIyS 120
DB 888 AATTGTGATTTTCTTTTACTGAGCTTCACACGTTACTGTAAATAATGAAGAAAG 947
QY 121 GluIySGInGIyIleGluIySGInIleIleuSerSerAlaIaAspIleIaIa 140
DB 948 AAACAAGAGAGAGATTGAGAAAGGCAAACTCTGCTTCAGCAGCAGCATTCCTGCC 1007
QY 141 ThrValGlnHisThrMetAlaCysHisIleuValIySArgThrHisArgAlaIleuPhe 160
DB 1008 ACAAGTACAGACACAATGGCATGTCATCTTGTGAAGAAACACATCGGCTATTCTGTTT 1067
QY 161 CysIySGInArgAspLeuLeuProGlnAsnAsnAlaValIleuValAlaSerGIyGIyVal 180
DB 1068 TGTAAAGCAGAGAGACTTGTTACCTCAAAATATGCAGTACGTGTCATCTGCTGCTGTC 1127
QY 181 AlaSerAspPheTyrlIeArgArgAlaIleuGlnIleuThrAsnAlaThrGlnCysThr 200
DB 1128 GCAAGTACTTCTATATCCGAGAGCTCGAATAATTTTAAACAACGCAACACAGTGCAC 1187
QY 201 LeuLeuCysProProPArgLeuCysThrAspAsnGIyIleMetIleAlaTrpAsnGIy 220
DB 1188 TTGTGTGCTCTCTCCAGCATATGACATGATATGATGATGATGATGATGATGATGAT 1247
QY 221 IleGluArgLeuArgAlaGIyLeuGIyIleuHisAspIleGIyIleArgTyrlu 240
DB 1248 ATGAAGAAGCTACGTGCTGCTGGCATTTTACATGACATAGAAGCATCGCTATGAA 1307
QY 241 ProIySerProLeuGIyValAspIleSerIySGInValGIyGlnAlaSerIleIySVal 260
DB 1308 CCAAAAGTCTCTTGGAGTAGACATATCAAAAGAGTGAAGAGCTTCATATAAAGTA 1367
QY 261 ProGlnLeuIySmetGIuile 267
DB 1368 CCAATTAATAAATGAGATTA 1388
RESULT 9
ADA52832
ID ADA52832 standard; cDNA; 2208 BP.
XX
AC ADA52832;
XX
DT 20-NOV-2003 (first entry)
XX
DE Human coding sequence, SEQ ID 400.
XX
CY Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;
KW Gene Therapy; human; secretory protein; membrane proteins; cancer;
KW Inflammatory disease; osteoporosis; neurological disease; gene; ss.
XX
OS Homo sapiens.
XX
PN BP1293569-A2.
XX
PD 19-MAR-2003.
XX
PF 21-MAR-2002; 2002EP-00006586.
XX
PR 14-SEP-2001; 2001JP-00328381.
XX
PR 24-JAN-2002; 2002US-0350435P.
XX
PA (HELI-) HELIX RES INST.
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
FI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Maeuho Y;

DR WPI; 2003-395539/38.
DR - P-PSDB; ADA54471.
XX
PT New polynucleotides encoding full-length polypeptides, e.g. secretory
PT and/or membrane proteins, useful for developing medicines for diseases in
PT which the gene is involved, or as target molecules for gene therapy.
XX
PS Claim 1; SEQ ID NO 400; 205bp; English.
XX
CC The present invention relates to novel human secretory or membrane
CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
CC ADA54071). The coding sequences are useful in the gene therapy of
CC diseases caused by abnormalities of the proteins, e.g. cancer,
CC inflammatory diseases, osteoporosis or neurological disease.
XX
SQ Sequence 2208 BP; 660 A; 454 C; 451 G; 643 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 3,3e-126 Length: 2208
Score: 1204.00 Matches: 239
Percent Similarity: 89.51% Conservative: 0
Best Local Similarity: 89.51% Mismatches: 4
Query Match: 86.93% Indels: 24
DB: 10 Gaps: 1

US-10-649-273-2_COPY_148_414 (1-267) x ADA52832 (1-2208)
QY 1 MetGluAlaHisAlaIleuThrIleArgLeuThrAsnIyValGIuPheProPheLeuVal 20
DB 785 ATGGAGGCTCAGACGCTTATCTATAGGTTGACCAAAATGAGAAATTTCTTTTATGTT 844
QY 21 LeuLeuIleSerGIyGIyHisCysLeuLeuAlaIleValGIyValSerAspPheLeu 40
DB 845 CTTTGTGATTTCTGGAGGTACGTCTGTTGGCATTTAGTTCAAGAGATTTCAAGATTTCTG 904
QY 41 LeuLeuGIyIySerLeuAspIleAlaProGIyAspMetLeuAspIyValAlaArgArg 60
DB 905 CTTTGGAAAGTCTTTGGACATGACACAGGAGCATGCTTGACAAAGGTGGCAAGAAG 964
QY 61 LeuSerLeuIleIyHisIProGluCysSerThrMetSerGIyGIyIyValIleGIuHis 80
DB 965 CTTCCTTAATTAACATCCAGAGTCTCACCATAGTGGTGGAAACCATAGAAACAT 1024
QY 81 LeuAlaIySGInGIyAaNaArgPheHisPheAspIleIySPProPLeuHisIaIaIyS 100
DB 1025 TTGGCCCAAGAGAAATAGATTTCATTTGACATCAAACTCCCTTCATCATCTAA 1084
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspIySIIleIleMetIyS 120
DB 1085 AATTGTGATTTTCTTTTACTGAGCTTCACACGTTACTGATTAATAATTAAGAAAG 1144
QY 121 GluIySGInGIyIleGluIySGInIleIleuSerSerAlaIaAspIleIaIa 140
DB 1145 GAAAAAGAGAGAGTATGAGAAAGGCGCAAACTCTGCTTCAGCAGCAGCATTCGCTGCC 1204
QY 141 ThrValGlnHisThrMetAlaCysHisIleuValIySArgThrHisArgAlaIleuPhe 160
DB 1205 ACAAGTACAGACACAATGGCATGTCATCTTGTGAAGAAACACATCGGCTATTCTGTTT 1264
QY 161 CysIySGInArgAspLeuLeuProGlnAsnAsnAlaValIleuValAlaSerGIyGIyVal 180
DB 1265 TGTAAAGCAGAGAGACTTGTTACCTCAAAATATGACATGCTGCTGCTGCTGCTGCTG 1324
QY 181 AlaSerAspPheTyrlIeArgArgAlaIleuGlnIleuThrAsnAlaThrGlnCysThr 200
DB 1325 GCAAGTAACTTCTGATCCGAGAGCTCGAAATTTTAAACAACGCAACACAGTGCAC 1384
QY 201 LeuLeuCysProProPArgLeuCysThrAspAsnGIyIleMetIleAlaTrpAsnGIy 220
DB 1385 TTGTGTGCTCTCTCCAGCATGACATGATTAATGGAATTAATGATGATGATGATGATG 1435
QY 221 IleGluArgLeuArgAlaGIyLeuGIyIleuHisAspIleGIyIleArgTyrlu 240

Db 1435 ----- 1435
Qy 241 ProlyscysProleuGlyValAsp11SerLysGlyValGlyGluAlaSer11LeuVal 260
Db 1436 ---TGATGTCCTCTTGAGTACATATCAAAAGAGTTGGAGAGCTTCCATAAAGTA 1492
Qy 261 ProGlnLeuLysMetGlu11Le 267
Db 1493 CCACAATTAATAATGAGAGATA 1513
RESULT 10
ADQ24627
ID ADQ24627 standard; DNA; 2890 BP.
XX
AC ADQ24627;
XX
DT 26-AUG-2004 (first entry)
XX
DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 7447.
XX
KM soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
KW de.
XX
OS Homo sapiens.
XX
PN MO2004048938-A2.
XX
PD 10-JUN-2004.
XX
PF 26-NOV-2003; 2003WO-US038193.
XX
PR 26-NOV-2002; 2002US-0429739P.
XX
PA (PROT-) PROTEIN DESIGN LABS INC.
XX
PI Aziz N, Ginsburg WM, Zlotnick A;
XX
DR WPI; 2004-441208/41.
XX
PT Early detection of soft tissue sarcoma comprises determining expression
PT of a gene in a first soft tissue sample and a normal soft tissue sample
PT and comparing the gene expression, also useful in treating soft tissue
PT sarcoma.
XX
XX
PS Example 2; SEQ ID NO 7447; 210bp; English.
XX
XX The invention relates to a novel method for detecting soft tissue sarcoma
CC which comprises obtaining a first soft tissue sample from an individual
CC and a normal soft tissue sample from the same or different individual,
CC determining the expression of a gene in both samples and comparing the
CC expression of the gene in both soft tissue samples, where a higher level
CC of protein expression in the first soft tissue sample indicates the
CC presence of soft tissue sarcoma. The method of the invention has
CC cytostatic applications and may be useful for detecting soft tissue
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC acid sequences may be useful in diagnostic and screening applications.
CC The current sequence is that of a human soft tissue sarcoma-upregulated
CC DNA of the invention. The current sequence is not shown within the
CC specification per se but was submitted in CD format by the inventor.
XX
SQ Sequence 2890 BP; 869 A; 609 C; 611 G; 789 T; 0 U; 12 Other;
Alignment Scores:
Pred. No.: 4.83e-126 Length: 2890
Score: 1204.00 Matches: 239
Percent Similarity: 89.51% Conservative: 0
Best Local Similarity: 89.51% Mismatches: 4
Query Match: 86.93% Indels: 24
DB: 12 Gaps: 1
US-10-649-273-2_COPY_148_414 (1-267) x ADQ24627 (1-2890)
Qy 1 MetGluAlaHisAlaLeuThr1LeArgLeuThrAlaLysValGluPheProPheLeuVal 20

Db 1442 ATGAGGCTCATGCACTTACTATTAAGTTGACCAATAAGATTTCTTTTATAGTT 1501
Qy 21 LeuLeu11SerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
Db 1502 CTTTGAATTTCTTGAGAGTACAGTCTGTTGGCATTAAGTTCAAGATTTTCTCG 1561
Qy 41 LeuLeuGlyLysSerLeuAsp11Lea1aProGlyAspMetLeuAspLysVal1aAArgArg 60
Db 1562 CTTCTTGGAAGTCTTTGGACATATGACACAGAGTGCATCTTGACAAAGGTGGCAAGACA 1621
Qy 61 LeuSerLeu11LeuHis1aProGlyLysSerThrMetSerGlyGlyVal1a11LeuHis 80
Db 1622 CTTCTTAATAATAAATCATCCAGAGTCTCCACATGAGTGGGGAAGCCTATGAACAT 1681
Qy 81 LeuAlaLysGlnGlyAsnArgPheHisPheAsp11LeuProProLeuHisHisAlaLys 100
Db 1682 TTGGCCAAACAAAGAAATGATTTCAATTTTGACATCAACCTCCCTTGACATCATGCTAA 1741
Qy 101 AsnCyAspPheSerPheThrGlyLeuGlnHisValThrAspLys11LeuMetLysLys 120
Db 1742 AATGTGATTTTCTTTTACTGAGCTTCAACAGTACTGATTAATAATAATGAAGAAAG 1801
Qy 121 GlnLysGlnGlnGly11LeuLysGlyGln11LeuSerSerAla1aAsp11Lea1a1a 140
Db 1802 GAAAAAGAGAAAGTAATGAGAAAGGCGCAAACTCTGCTTCAGACAGCATTTGCTGCC 1861
Qy 141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAla1LeuPhe 160
Db 1862 ACAGTACGACACACAAATGACATGTCATCTTGGAAGAAGACATGCGCTATTCGTTT 1921
Qy 161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuVal1aSerGlyGlyVal 180
Db 1922 TGTATGTCCTCTCCCTCCAGACTATGACTGATTAAGGCAATTAATGATGCA----- 2092
Qy 181 AlaSerAsnPheTyrl1eArgArgAlaLeuGlu11LeuThrAsnAlaThrGlnCysThr 200
Db 1982 GCAGTAATCTTGATCCGACAGAGCTCGAAATTTTAACAAAACCAACACAGTGCACAT 2041
Qy 201 LeuLeuCysProProProArgLeuCysThrAspAsnGly11LeuMet11Lea1aTrpAsnGly 220
Db 2042 TTGTTGTGTCCTCTCCCTCCAGACTATGACTGATTAAGGCAATTAATGATGCA----- 2092
Qy 221 IleGluArgLeuArg1aGlyLeuGly11LeuHisAsp11LeuGly11LeaArgTyGlu 240
Db 2092 ----- 2092
Qy 241 ProLysCysProleuGlyValAsp11SerLysGlyValGlyGluAlaSer11LeuVal 260
Db 2093 ---TGATGTCCTCTTGAGTACATATCAAAAGAGTTGGAGAGCTTCCATAAAGTA 2149
Qy 261 ProGlnLeuLysMetGlu11Le 267
Db 2150 CCACAATTAATAATGAGAGATA 2170
RESULT 11
ADE31345/c
ID ADE31345 standard; DNA; 3358 BP.
XX
AC ADE31345;
XX
DT 29-JAN-2004 (first entry)
XX
DE Human diagnostic and therapeutic polynucleotide (dthp), SEQ ID NO 100.
XX
XX diagnostic and therapeutic polynucleotide; dthp; antiarteriosclerotic;
XX antiinflammatory; cerebroprotective; antiipsemic; antidiabetic;
XX immunosuppressive; neuroprotective; antitropic; neuroleptic; tranquilizer;
XX osteopathic; antirheumatic; antirheumatic; cytostatic; hepatotropic;
XX vitaminic; haemostatic; anti-HIV; antithyroid; thyronumetic;
XX dermatological; antibacterial; fungicide; antiparasitic; anticonvulsant;
XX thrombolytic; anticoagulant; anorectic; vasotropic; antidiuretic;
XX gene therapy; protein replacement therapy; human; gene; ds-

XX Homo sapiens.
OS
XX
XX WO2003062376-A2.
PN
XX
XX 31-JUL-2003.
PD
XX
XX 13-JAN-2003; 2003WO-US001096.
PF
XX
XX 16-JAN-2002; 2002US-0349384P.
PR
XX
XX 17-JAN-2002; 2002US-0349413P.
PR
XX
XX 17-JAN-2002; 2002US-0349946P.
XX
XX
XX (INCY-) INCYTE GENOMICS INC.
PA
XX
XX Jones AL, Dahl CR, Gietzen D, Chim J, Dufour GE, Jackson JL,
PI
XX Yu JY, Tusson O, Yap PE, Ameshey SR, Dam TC, Liu TF, Gerstin EH,
PI
XX Peralta CH, Lewis SA, Chen A, Marwaha R, Lan RY, Uraishka ME,
PI
XX Kristnam SR, Kolluru V, Panesar IS;
XX
XX WPI; 2003-636732/60.
DR
XX P-PSDB; ADE31156.
XX
XX
XX New human diagnostic and therapeutic polynucleotides and polypeptides,
PT
XX useful for diagnosing, treating or preventing e.g. leukemia, brain
PT
XX cancer, atherosclerosis, AIDS, thyroiditis, infections, obesity, stroke
PT
XX or Alzheimer's.
XX
XX
XX Claim 1; SEQ ID NO 100; 634pp; English.
PS
XX
XX
XX The invention relates to a novel isolated human diagnostic and
CC
XX therapeutic polynucleotide (designated ditp). The novel ditp
CC
XX polynucleotide comprises: any of 188 DNA sequences consisting of 195-7798
CC
XX base pairs fully defined in the specification; a polynucleotide
CC
XX comprising a naturally occurring polynucleotide sequence at least 90%
CC
XX identical to the ditp polynucleotide; a polynucleotide complementary to
CC
XX the ditp polynucleotide or its polynucleotide which is at least 90%
CC
XX identical; or an RNA equivalent of any of the polynucleotides mentioned
CC
XX above. The ditp polynucleotides have the following activities:
CC
XX antiarteriosclerotic, antiinflammatory, cerebroprotective, antilipemic,
CC
XX antidiabetic, immunosuppressive, neuroprotective, nootropic, neuroleptic,
CC
XX tranquilizer, osteopathic, antirheumatic, antirheumatic, cytostatic,
CC
XX hepatotropic, vitronic, haemostatic, anti-HIV, antithyroid, thyromimetic,
CC
XX dermatological, antibacterial, fungicide, antiparasitic, anticonvulsant,
CC
XX thrombolytic, anticoagulant, anorectic, vasotropic, and antitumor. The
CC
XX novel DITP polynucleotides polypeptide can be used in gene therapy and
CC
XX protein replacement therapy. The ditp polynucleotides or DITP
CC
XX polypeptides are useful for diagnosing, preventing or treating diseases
CC
XX associated with the expression of human molecules. In particular, these
CC
XX diseases include cancers (e.g. adenocarcinoma, leukemia, melanoma, brain
CC
XX cancer, breast cancer, cervix cancer, bone cancer, liver cancer, lung
CC
XX cancer) or other cell proliferative disorders (e.g. arteriosclerosis,
CC
XX atherosclerosis, cirrhosis, hepatitis, polycythemia vera, primary
CC
XX thrombocytopenia), autoimmune/inflammatory disorders (e.g. AIDS,
CC
XX Addison's disease, thyroiditis, Crohn's disease, Graves' disease,
CC
XX Hashimoto's thyroiditis, multiple sclerosis, osteoarthritis, rheumatoid
CC
XX arthritis, scleroderma, systemic lupus erythematosus), infections (e.g.
CC
XX viral, bacterial, fungal or parasitic infection), developmental disorders
CC
XX (e.g. Cushing's syndrome or epilepsy), endocrine disorders (e.g.
CC
XX thombosis, hypopituitarism, hypogonadism, gigantism, goiter) metabolic
CC
XX disorders (e.g. hypercholesterolaemia, hypoglycaemia, diabetes,
CC
XX hyperlipidaemia, obesity), neurological disorders (e.g. ischemic
CC
XX cerebrovascular disease, stroke, Alzheimer's disease, Pick's disease,
CC
XX Huntington's disease, Parkinson's disease, Creutzfeldt-Jakob disease,
CC
XX anxiety, schizophrenia), gastrointestinal disorders (e.g. ulcers,
CC
XX transport disorders (e.g. akinesia or multidrug resistance), or
CC
XX connective tissue disorders (e.g. Paget's disease or rickets). This
CC
XX polynucleotide sequence represents one of the human ditp DNA sequences
XX
XX of the invention.

Pred. No.:	2,99e-102	Length:	3358
Score:	995.50	Matches:	250
Percent Similarity:	44.82	Conservative:	1
Best Local Similarity:	44.64	Mismatches:	7
Query Match:	71.88	Indels:	307
DB:	10	Gaps:	2

US-10-649-273-2 COPY 148 414 (1-267) X ADE31345 (1-3358)

```

10 LeuThrValGluPheProPheValLeuLeuLeuSerGlyGlnGlyLeu 29
:::

```

Db 3067 GTGACCAATAAGTAGAATTCCTTTTCTAGTTCT-TTGATTTCTGGAGGTCACCTGCTG 3009

QY 30 LeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlnGlyLysSerLeuAspIleAla 49

Db 3008 TGGCATTAGTTCAGGAGTTTCAGATTTCTGCTTCTTGGAAGTC-TTGACATAGCA 2950

50 Proglyaspmet----- 53

Db 2949 CCAGGTGACATGCGTTGACCAAGTAATTAACTTCTCCATTCTTTTGTTATGT 2890

Q_Y 53

Db 2889 TGTCAATTCACCTAAGTAGCAATAGATGTGCTACCAACCATTCACCTAATATTTCTGAA 2830

Qy 53 ----- 53

Db 2829 TTTTATCTTAGTAACAGACACAAATTCACATATGCTGAGAAAAATAGAAAGCAGTA 2770

Q7 ----- 53

Db 2769 GTACACGAATTATATTCTTAGCCTTTCTTAATAAATGTAAGAGGTCATATCTGTA 2710

53 ----- 53

Db 2709 CATAAAGGCTGAATAGTTTGCAGATACAGTTATGTATTTGCCAATAATGTATGTGA 26

$\alpha\gamma$	53	53
----------------	----	----

Db 2649 AGAACGTCCTTCGTAACATACTGCAAAAAGGTAATAAGAGATATATAGA 2590

α_3 ----- 53

Db 2589 TTACATAAGGACATTAAAGATGCAATGCACAGATTAAATCACACAATTACTTACACCA 25

$\alpha\gamma$	53
-----	53

Db 2529 CAGACAGGTCCTCCCGCACCCCTTGTGTTTAGAATACTACAGAGGCTACTGCCATAT 2470

α_3 53

Db 2469 ATAGGA ACTACAACAACGACGTCCTCCACAGTGAATAATAGGAAGTATAGG 2410

α_3 ----- α_3

Db 2409 ACAAGTCTTATTATGACGTCATCATTAAGCAGTTATTGTCAACTTCAAGCCATT 2350

α_3 ----- α_3

Db 2349 CCAACCAATAGAGAGCAACATAGACAGGGCAGTGTATTTGTTGGGTC 2290

α 53 ----- 53

Db 2289 ATCATAAGGACAGGTTGTCTGCTTACCTGAATATCAGCTATAGTCTATATTTGCCAA 2230

53 ----- 53

Db 2229 GTATAGCATGTTTATTCAATCAGGCGTTTTTGTGTTGTTAGTAATTTCAATTTATTT 2170

54 -----leu 54

Db 2169 CCTTGCATCTTTGCTTCACAGTATTAAATTTATGACTCTAAAAAATATGTTCTTT 2110

Alignment Scores:

QY 55 AsplvVala1aAargLeuSerleu1leuYsh1sPProGluCySeSerThMeSeSergly 74
DB 2109 GATAG-GTGGCAAGAGACTTTCTTATATAAATCAGAGTGTCCACCATGAGTGT 2051
QY 75 -GlyVala1a1eGluH1sleuA1a1eGluG1yAasnArgPheH1sPheApl1e1yPr 94
DB 2050 GGGGAAAGCCATAGAACTTTGGCCAAACAGAAATGATTTCAATTTGACATCAACC 1991
QY 94 oProLeuH1sH1sAla1yAasnCyAspPheSerPheThrg1yLeuG1nH1sVal1ThraS 114
DB 1990 TCCCTTGATCATGCTAATAAATGTGATTTTCTTTACTGACCTTCAACGCTTACTCA 1931
QY 114 pLyS1le1leuMetLySgluYsGluGlu-----Gly1leGluYsG 124
DB 1930 TAAATATATATGAAAAAGAAAAAGAGATATTTCTAATTAGTAAAGTGAACA 1871
QY 125 -----Gly1leGluYsG 129
DB 1870 GATTAATATTCCTGATTTGCTTAAATAATAGCTGCTCATTTCTGACGATATTGAAGG 1811
QY 129 lYg1n1leuSeSer1a1a1aAsp1le1a1a1eThrVal1G1nH1sThMeAlaCySH 149
DB 1810 GGCATAATCTCTCTTACAGACAGACATTTGCTGCCACAGTACAGACACAAATGCAATGTC 1751
QY 149 lSleuVal1yAArgThrH1sArgAla1le1eUphCyS1ySg1naRgaSpleuLeuProG 169
DB 1750 ATCTTGTAAGAAAGACACATCGGGCTATTCGTTTGTAGACAGAGACTTGTTACCTC 1691
QY 169 lnaAsnaH1aVal1eUVal1a1aSerG1yG1yVal1a1aSerAsnPhery1leArgH1ga 189
DB 1690 AAAATAATAGCAGTACTGCTGATCTGT-GGTGTGCAAGTAACTTATATATATCCACAGAG 1632
QY 189 lalauGlu1leuThraSna1a1Thrg1nCySeThleuLeuCySProP-ROpAArgLeuC 209
DB 1631 CTCTGAAATTTTAC-AACGCAACACAGTGCATTTGTGTCTCTCCCAACATAT 1573
QY 209 ySThrAspAsnG1y1leMet1leA1aTPaenG1y1leGluArgLeuA1aG1yLeuG 229
DB 1572 GCACTGATATATGCTTATGATTCATGAAATGATTAAGAACTACCTGCTGCTTGG 1513
QY 229 lY1leUuH1aAsp1leGluG1y1leArgY1yGluProLyS1yAspProLeuG1yVal1Asp1 249
DB 1512 GCATTTTACATGACATGAGAGGCAATCCGCTATGAACCAAAATGTCTCTTGAGATAGACA 1453
QY 249 lSeSerLySgluVal1G1yUu1a1aSer1le1ySvalProG1nleuYsMetGlu1le 267
DB 1452 TATCAAAAGAAAGTTGAGAGGCTTCATTAAGTACCAAAATTAATAAGAGATA 1397
RESULT 12
ABQ75508 standard; DNA; 1572 BP.
XX
AC ABQ75508;
XX
DT 07-NOV-2002 (first entry)
XX
DE Murine sialoglycoprotease-like gene sequence SEQ ID NO:7.
XX
KW Murine; mouse; protease; calcium activated neutral protease type 5;
KW CAPN5; tryptase 4; sialoglycoprotease; enzyme; genetic disease;
KW neurological; neuropsychological; psychotic illness; transgenic animal;
KW gene; ds.
XX
OS Mus musculus.
XX
PN W0200245491-A2.
XX
PD 13-JUN-2002.
XX
XX 05-DEC-2001; 2001WO-US046405.
XX
XX 06-DEC-2000; 2000US-0251803P.
XX
PR 06-DEC-2000; 2000US-0251820P.
PR

PR 13-DEC-2000; 2000US-0255971P.
XX
XX (DELT-) DELTAGEN INC.
XX
PI Allen KD, Leviten MW;
XX
DR WPI; 2002-657389/70.
XX
PT Novel transgenic animal, comprising a disruption in protease target gene,
PT is useful for identifying agents that ameliorates a phenotype associated
PT with a disruption in a protease target gene.
XX
PS Example 3; Fig 7; 62pp; English.
XX
CC The present invention describes a non-human transgenic animal (I)
CC comprising a disruption in a protease target gene (PG) selected from
CC calcium activated neutral protease type 5 (CAPN5) gene, tryptase 4 gene
CC and sialoglycoprotease-like gene. Also described is a targeting construct
CC (II), comprising a first polynucleotide sequence homologous to at least a
CC first portion of PG, a second polynucleotide sequence homologous to at
CC least a second portion of PG and a selectable marker. (II) is useful for
CC producing a transgenic mouse comprising a disruption in a protease target
CC gene, by introducing (II) into a cell, introducing the cell into a
CC blastocyst, implanting the resulting blastocyst into a pseudopregnant
CC mouse, where the pseudopregnant mouse gives birth to a chimeric mouse,
CC and breeding the chimeric mouse to produce the transgenic mouse. (I) is
CC useful for identifying an agent that modulates the expression or function
CC of a protease target gene, by administering an agent to (I) and
CC determining whether the expression or function of the disrupted protease
CC target gene in (I) is modulated. (I) is also useful for testing the
CC efficacy of proposed genetic and pharmacological therapies for human
CC genetic diseases, such as neurological, neuropsychological or psychotic
CC illness. The present sequence represents murine sialoglycoprotease-like
CC gene sequence, which is used in an example from the present invention
XX
SQ Sequence 1572 BP; 459 A; 337 C; 340 G; 429 T; 0 U; 7 Other;
XX
Alignment Scores:
Pred. No.: 2, 67e-101 Length: 1572
Score: 993.00 Matches: 210
Percent Similarity: 84.21% Conservative: 14
Best Local Similarity: 78.95% Mismatches: 38
Query Match: 70.97% Indels: 6
DB: Gaps: 2
US-10-649-273-2_COPY_148_414 (1-267) x ABQ75508 (1-1572)
QY 1 MetGluA1aH1a1a1eUThr1leA1a1a1eUThraSna1yA1a1eUphProPheUVal 20
DB 445 ATGAGAGGCTCAGCACTGACTATTAGGCTCACCAATRAAGTAGATTN-CCTTTTAAAGTT 503
QY 21 leuLeu1leSeSerG1yG1yH1sCySleuLeuA1a1eUVal1G1nG1yVal1a1eSerAspPheLeu 40
DB 504 CTTTGTGATTTCTGGCGGCACTGCCCTGTGAKMTTGTCCAAAGTNTTTCCGATTTCTCTG 563
QY 41 leuLeuG1ySeSer1eUAsp1leA1aPProG1yAspMet1eUAsp1ySval1a1aArgArg 60
DB 564 CTCCTTGGGAAGCTTTGACATATAGACACAGGCAATGCTTACAAAGTGGCAAGAGA 623
QY 61 leuSer1eU1leYsh1sPProGluCySeSerThMeSeSerg1yG1ySval1a1eGluH1s 80
DB 624 CTTTCTTATCAAAATCATCAGAAATTTCTACAAATAGAGGTGGAAGAGTATAGAACAG 683
QY 81 leuA1a1ySgluG1yAasnArgPheH1sPheApl1e1ySProPProLeuH1sH1a1a1yS 100
DB 684 TTGGCAAAAGACGAAATGATTCATTTACTATATCAACTATGCAAGAAATGCTAAG 743
QY 101 AsnCysAspPheSerPheThrg1yLeuG1nH1sVal1ThraSple1le1leMetLyS1yS 120
DB 744 AATTCGATTTTCTTTCACGGGACTTCAACATATTAATGATTAATTAACACACAG 803
QY 121 GluLySgluG1yG1y1leGluYsG1yG1n1leUSeSer1a1a1aAsp1leA1a1a 140

AC ADL86725;
 XX 20-MAY-2004 (first entry)
 XX
 XX DNA up-regulated in murine multipotent progenitor cells Segid 3118.
 DE
 XX gene potential; multi-lineage; cell commitment; haematopoietic stem cell;
 KM HSC; multipotent progenitor; MPP; common lymphoid progenitor; CLP;
 KM common myeloid progenitor; CMP; bone marrow stem cell; mouse; murine; de.
 XX
 OS Mus sp.
 XX WO2003093445-A2.
 XX
 XX 13-NOV-2003.
 PD
 XX 05-MAY-2003; 2003WO-US014114.
 PF
 XX 03-MAY-2002; 2002US-0377383P.
 PR
 XX (STOW-) STOWERS INST MEDICAL RES.
 PA
 XX Li L;
 PI
 XX WPI; 2004-022656/02.
 DR
 XX
 XX Classifying an unknown multi-lineage affiliated gene comprises isolating
 PT expressed nucleic acid sequences from the discrete cell sub-populations.
 PS
 XX Claim 8; SEQ ID NO 3118; 123pp; English.
 XX
 XX This invention relates to a novel method for predicting gene potential by
 CC associating nucleic acid sequences of unknown function with particular
 CC sub-population profiles. Specifically, it refers to classifying an
 CC unknown multi-lineage affiliated gene by collecting hybridisation data to
 CC develop a gene expression map, in order to determine the discrete sub-
 CC population where it is expressed. The present invention describes methods
 CC for predicting the lineage commitment of genes associated with the self-
 CC renewing haematopoietic (blood) stem cells (HSCs), as well as the non-
 CC self renewing multipotent progenitors (MPPs), common lymphoid progenitors
 CC (CLPs) and common myeloid progenitors (CMPs), which are collectively
 CC referred to as bone marrow stem cells populations. As such, these methods
 CC can be used to identify associated multi-lineage affiliated genes and
 CC hence the underlying molecular mechanisms in physiological haematopoietic
 CC development. This polynucleotide sequence is DNA associated with a murine
 CC MPP sub population of cells of the invention.
 CC
 XX
 XX Sequence 371 BP; 104 A; 75 C; 84 G; 92 T; 0 U; 16 Other;
 SQ
 XX
 XX Alignment Scores:
 Pred. No.: 1.18e-43 Length: 371
 Score: 468.00 Matches: 92
 Percent Similarity: 78.86% Conservative: 5
 Best Local Similarity: 74.80% Mismatches: 26
 Query Match: 33.79% Indels: 0
 DB: 12 Gaps: 0
 US-10-649-273-2_COPY_148_414 (1-267) x ADL86725 (1-371)
 QY 137 AspllelaaiaatnrvalgnhieshrmecaAcyshisileuvallyeargthrhisarg 156
 DB 2 GACATTCGCTGCTGGGATACACATGCAACAGCGTCCACCTTGCGCAAAAGAACACATGCC 61
 QY 157 AlaileuPheCyLySGlnarGAspleuLeuProGlnasnaAlaValleuValala 176
 DB 62 GCTATTCTGTTTGCACGACGAAATTTGCTCTCCAGCTAACGCAGTATTAGTTGTA 121
 QY 177 SerGlyGlyValAlaAserAspheYrIleargAgaIaleuGlnIleuThrAsnaA 196
 DB 122 TCTGAGAGGTGTGCAAGTAACCTTGTAACATCCGAAAGACATTGGAATGTCGCCAATGCA 181
 QY 197 ThrGlnCysThrIleuLeuCySProProArgLeuCySthAspAsnGlyIleuGctle 216

DB 182 ACGCAGTCACGTTGTGTGTCACCTCCAGACCTGTGCATGCAACATGGCATGATGATT 241
 QY 217 AlaTTPAsnglyIleGluarGleuArGalaGlyleuGlyIleuwhisaspIleGlyugly 236
 DB 242 GCATGGAATGCAATTGGAATGAGATTACGTGCGNGCTTNGGCTNTTACNTGANNATGAGAGAC 301
 QY 237 ILeargYrGlnProIyCySProleuGlyValAspIleSerlyGluValGlyIuAla 256
 DB 302 ATCCNNATATGNAACCAAAATNTGNTVTGAGTGAACATMTCCAGAGAGTTCGCAAGCT 361
 QY 257 SerIleIys 259
 DB 362 GCCATAAAA 370
 RESULT 15
 ADL86726
 ID ADL86726 standard; DNA; 371 BP.
 XX
 XX ADL86726;
 XX
 XX 20-MAY-2004 (first entry)
 DT
 XX
 XX DNA up-regulated in murine multipotent progenitor cells Segid 3119.
 DE
 XX gene potential; multi-lineage; cell commitment; haematopoietic stem cell;
 KM HSC; multipotent progenitor; MPP; common lymphoid progenitor; CLP;
 KM common myeloid progenitor; CMP; bone marrow stem cell; mouse; murine; de.
 XX
 OS Mus sp.
 XX WO2003093445-A2.
 XX
 XX 13-NOV-2003.
 PD
 XX 05-MAY-2003; 2003WO-US014114.
 PF
 XX 03-MAY-2002; 2002US-0377383P.
 PR
 XX (STOW-) STOWERS INST MEDICAL RES.
 PA
 XX Li L;
 PI
 XX WPI; 2004-022656/02.
 DR
 XX
 XX Classifying an unknown multi-lineage affiliated gene comprises isolating
 PT expressed nucleic acid sequences from the discrete cell sub-populations.
 PS
 XX Claim 8; SEQ ID NO 3119; 123pp; English.
 XX
 XX This invention relates to a novel method for predicting gene potential by
 CC associating nucleic acid sequences of unknown function with particular
 CC sub-population profiles. Specifically, it refers to classifying an
 CC unknown multi-lineage affiliated gene by collecting hybridisation data to
 CC develop a gene expression map, in order to determine the discrete sub-
 CC population where it is expressed. The present invention describes methods
 CC for predicting the lineage commitment of genes associated with the self-
 CC renewing haematopoietic (blood) stem cells (HSCs), as well as the non-
 CC self renewing multipotent progenitors (MPPs), common lymphoid progenitors
 CC (CLPs) and common myeloid progenitors (CMPs), which are collectively
 CC referred to as bone marrow stem cells populations. As such, these methods
 CC can be used to identify associated multi-lineage affiliated genes and
 CC hence the underlying molecular mechanisms in physiological haematopoietic
 CC development. This polynucleotide sequence is DNA associated with a murine
 CC MPP sub population of cells of the invention.
 CC
 XX
 XX Sequence 371 BP; 104 A; 75 C; 84 G; 92 T; 0 U; 16 Other;
 SQ
 XX
 XX Alignment Scores:
 Pred. No.: 1.18e-43 Length: 371
 Score: 468.00 Matches: 92
 Percent Similarity: 78.86% Conservative: 5
 Best Local Similarity: 74.80% Mismatches: 26
 Query Match: 33.79% Indels: 0

DB: 12 Gaps: 0

US-10-649-273-2_COPY_148_414 (1-267) x ADL86726 (1-371)

```

Qy 137 AsplleAlaIATrrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArg 156
    |||
Db 2 GACATTGCTGCTGCGGTACAGCATCAACGCGTGCACCTTGCGAAMAGAACACTCGC 61
    |||

Qy 157 AlaIleuPheCysLysGlnArgAspleuLeuProGlnAsnAsnAlaValLeuValAla 176
    |||
Db 62 GCTATTCTGTTTGGAAAGCAGAAATTGGCTCTCCAGCTAACGCAATTAAGTTGTA 121
    |||

Qy 177 SerGlyGlyValAlaSerAsnPheTyrIleArgArgAlaLeuGluIleLeuThrAsnAla 196
    |||
Db 122 TCTGAGGCTGTGCAAGTAACCTTGACATCCGAAAGCATTGGAAATTGTCGCAATGCA 181
    |||

Qy 197 ThrGlnCysThrLeuLeuCysProProArgLeuCysThrAspAsnGlyIleMetIle 216
    |||
Db 182 ACGCAGTGCACGTTGTGTGTCACCTCCAGACTGTGCATGACATGCAATGCAATGATT 241
    |||

Qy 217 AlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGly 236
    |||
Db 242 GCATGGAATGGAATTGAANGATTACGTGCNGCTTNGCCTTNTTACNTGANTAGAAAGAC 301
    |||

Qy 237 IleArgTyrGluProLysCysProLeuGlyValAspIleSerLysGluValGlyAla 256
    |||
Db 302 ATCCNMTATGNACCAGAAATNTCNTNTTGGAGTAGACATNTCCAGAGAAAGTTGCAGAGCT 361
    |||

Qy 257 SerIleLys 259
    |||
Db 362 GCCATATAAA 370
    |||

```

Search completed: November 10, 2005, 17:34:13
 Job time : 611.911 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: November 10, 2005, 16:52:13 ; Search time 5267.16 Seconds
(without alignments)
2456.267 Million cell updates/sec

Title: US-10-649-273-2_COPY_148_414
Perfect score: 1385
Sequence: 1 MEHALTRITRNKVEPEFLV.....DISKVGESIKVPOLKMEI 267

Scoring table: BL0SUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+ p2n.model -DEV=xlp
-Q/cgnt2_1/USFTO.spool_p/US10649273/runat_02112005_091338_15550/app.query.fasta_1.1429
-DB=GenEmbl -OFMT=faetap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pro -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10649273 @CGN 1.1.7442 @runat_02112005_091338_15550 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*
1: gb_ba.*
2: gb_hlg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	1385	100.0	1416	AR541929	Sequence
2	1385	100.0	1526	AR428809	Sequence
3	1385	100.0	1908	BC011904	Homo sapi
4	1385	100.0	2197	AR428803	Sequence

5	1358.5	98.1	1387	6	AR428808	Sequence
6	1358.5	98.1	1387	9	HS295148	Homo sapi
7	1348	97.3	1345	6	AX664697	Sequence
8	1348	97.3	1345	6	AX664695	Sequence
9	1231	1820	1844	6	BC058172	Mus muscu
10	1224	88.4	1017	10	BC038910	Mus muscu
11	1204	86.9	2208	6	AX713716	Sequence
12	1204	86.9	2208	9	AK055441	Homo sapi
13	1116	1546	1546	10	BC078974	Rattus no
14	950.5	68.6	1109	10	BC051211	Mus muscu
15	939	67.7	1522	5	BX934991	Gallus ga
16	938	67.7	860	5	BX930694	Gallus ga
17	934	67.4	1558	5	BX930963	Gallus ga
18	831.5	60.0	84115	9	AC013468	Homo sapi
19	824.5	59.5	14364	6	AR428807	Sequence
20	666	48.1	249601	2	AC114153	Rattus no
21	666	48.1	308652	2	AC121478	Rattus no
22	652.5	47.1	256751	10	AC122825	Mus muscu
23	610	44.0	875	6	CQ721898	Sequence
24	488.5	35.3	1474	3	AK113378	Ciona int
25	477.5	34.5	121251	5	AL591593	Zebrafish
26	435	31.4	117322	5	AL672217	Zebrafish
27	415.5	30.0	1576	3	AY051882	Drosophila
28	415.5	30.0	1601	6	CQ606432	Sequence
29	415.5	30.0	3656	6	CQ606431	Sequence
30	415.5	30.0	14679	2	AC018262	Drosophila
31	415.5	30.0	180263	3	AC010671	Drosophila
32	415.5	30.0	207432	3	AE003513	Drosophila
33	408	29.5	1672	8	AK070912	Oryza sat
34	407.5	29.4	1443	8	AY024338	Arabidops
35	407.5	29.4	1474	8	AY117283	Arabidops
36	407.5	29.4	1567	8	AY063864	Arabidops
37	401.5	29.0	1557	8	AY084577	Arabidops
38	389	28.1	110000	2	AP006501	Continuation (9 of
39	369	26.6	260271	1	AE017258	AB01258 Molbacteria
40	356.5	25.7	333800	1	SM551992	Al591792 Mus muscu
41	356	25.7	20021	10	AF367967	Mus muscu
42	356	25.7	179252	10	AF131205	Mus muscu
43	352	25.4	349116	1	AP003003	Mesozhizo
44	349	25.2	110000	1	AB017197	Continuation (2 of
45	342	24.7	756	6	AX685201	Sequence

ALIGNMENTS

RESULT 1
AR541929 1416 bp DNA linear PAT 08-OCT-2004
LOCUS AR541929 Sequence 177 from patent US 6743619.
DEFINITION AR541929
ACCESSION AR541929
VERSION AR541929.1 GI:53934009
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1416)
AUTHORS Tang,Y.T., Zhou,P., Goodrich,R., Liu,C., Asundi,V., Ren,F.,
Zhang,D., Zhao,Q.A., Yang,Y., Xue,A.D., Wehrman,T., Wang,J.-R.,
Wang,D. and Drmanac,R.T.
TITLE Nucleic acids and polypeptides
JOURNAL Patent: US 6743619-A 177 01-JUN-2004;
FEATURES
source location/Qualifiers
1..1416
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN

Alignment Scores:
Pred. No.: 1.37e-126 Length: 1416
Score: 1385.00 Matches: 267
Percent Similarity: 100.00% Conservatve: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0

DB: 6 Gaps: 0
US-10-649-273-2_COPY_148_414 (1-267) x AR541929 (1-1416)
QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20
DB 502 ATGGAGGCTCATGCACTTACTATTAGGTGAGCAATAAGATGAATTTCTTTTATTAGTT 561
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
DB 562 CTTTGTGATTTTGTGAGGTCACCTGCTGTGGCATTAGTTCAGAGAGTTTCAGATTTTCTG 621
QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60
DB 622 CTTCTTGAAAGTCTTTGGACATAGCACAGGTGACATGCTTGACAGAGTGGCAAGAGA 681
QY 61 LeuSerLeuIleLysHisProGlyLysSerThrMetSerGlyGlyLysAlaIleGluHis 80
DB 682 CTTTCTTTAATAAACAATCCAGAGTCTCCACCATGAGTGGGAAAGCCATGAAACAT 741
QY 81 LeuAlaLysGlnGlyAsnArgPheHisAspAspIleLysProProLeuHisHisAlaLys 100
DB 742 TTGGCCAAACAGAAATAGATTTCATTTTGACATCAAACTCCCTTGCAATGCTTAA 801
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValIleAspLysIleIleMetLysLys 120
DB 802 AATTGTGATTTTCTTTACTGGACTTCAACCGTTACGTAAATAATAATATGAAAAAG 861
QY 121 GluLysGluGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaAla 140
DB 862 GAAAAAGAGGAAGATTGAGAGAGGGCAAAATCCTGCTTCAGACAGACATTCCTGCC 921
QY 141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160
DB 922 ACAGTACAGCACACATGCGCATGTCATCTTGAAAAAGAACACATCGGCGTATTCGTTT 981
QY 161 CysLysGlnAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyVal 180
DB 982 TGTAAAGCAGAGACTTGTACTCAAAATATATGACATGCTGTTGCATCTGCTGCTGTC 1041
QY 181 AlaSerAsnPheTyrlleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThr 200
DB 1042 GCAAGTAACTTCTATATCCGAGAGCTCTGGAATTTTAAACAACGCAACACAGTGCAC 1101
QY 201 LeuLeuCySPProProProArgLeuCySPThrAspAsnGlyIleMetIleAlaIleAsnGly 220
DB 1102 TTGTGTGTCTCTCCCTCCAGACTATGCACTGATTAATGCAATTAATGCAATGCAATGCT 1161
QY 221 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyTyrlleArgTyrglu 240
DB 1162 ATTGAAGACTACGTGCTGGCTTGAGCATTTTACATGACATGAAAGGCATCCGCTATGAA 1221
QY 241 ProLysCySPProLeuGlyValAspIleSerLysGluValGlyGluAlaSerIleLysVal 260
DB 1222 CCAAAATGCTCTTGGAGTAGACATATCAAAAGAAAGTTGGAAGGCTTCCATTAAGTA 1281
QY 261 ProGlnLeuLysMetGluIle 267
DB 1282 CCACAAATTAATAATGAGATA 1302
RESULT 2
AR428809 1526 bp DNA linear PAT 18-DEC-2003
LOCUS AR428809
DEFINITION Sequence 23 from patent US 6642041.
ACCESSION AR428809
VERSION AR428809.1 GI:40188595
KEYWORDS
SOURCE unknown.
ORGANISM unknown.
REFERENCE 1 (bases 1 to 1526)
AUTHORS Chen,J., Feder,J.N., Nelson,T.C., Krystek,S.R. and Duclos,F.
TITLE Polynucleotides encoding a novel metalloprotease, MP-1

JOURNAL Patent: US 6642041-A 23 04-NOV-2003;
FEATURES Location/Qualifiers
source 1..1526
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Alignment Scores:
Pred. No.: 1..5e-126 Length: 1526
Score: 1385.00 Matches: 267
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-10-649-273-2_COPY_148_414 (1-267) x AR428809 (1-1526)
QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20
DB 1 ATGGAGGCTCATGCACTTACTATTAGGTGAGCAATAAGATGAATTTCTTTTATTAGTT 60
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
DB 61 CTTTGTGATTTTGTGAGGTCACCTGCTGTGGCATTAGTTCAGAGAGTTTCAGATTTTCTG 120
QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60
DB 121 CTTCTTGAAAGTCTTTGGACATAGCACAGGTGACATGCTTGACAGAGTGGCAAGAGA 180
QY 61 LeuSerLeuIleLysHisProGlyLysSerThrMetSerGlyGlyLysAlaIleGluHis 80
DB 181 CTTTCTTTAATAAACAATCCAGAGTCTCCACCATGAGTGGGAAAGCCATGAAACAT 240
QY 81 LeuAlaLysGlnGlyAsnArgPheHisAspAspIleLysProProLeuHisHisAlaLys 100
DB 241 TTGGCCAAACAGAAATAGATTTCATTTGACATCAAACTCCCTTGCAATGCTTAA 300
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValIleAspLysIleIleMetLysLys 120
DB 301 AATTGTGATTTTCTTTACTGGACTTCAACCGTTACGTAAATAATAATATGAAAAAG 360
QY 121 GluLysGluGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaAla 140
DB 361 GAAAAAGAGGAAGATTGAGAGAGGGCAAAATCCTGCTTCAGACAGACATTCCTGCC 420
QY 141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160
DB 421 ACAGTACAGCACACATGCGCATGTCATCTTGAAAAAGAACACATCGGCGTATTCGTTT 480
QY 161 CysLysGlnAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyVal 180
DB 481 TGTAAAGCAGAGACTTGTACTCAAAATATATGACATGCTGTTGCATCTGCTGCTGTC 540
QY 181 AlaSerAsnPheTyrlleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThr 200
DB 541 GCAAGTAACTTCTATATCCGAGAGCTCTGGAATTTTAAACAACGCAACACAGTGCAC 600
QY 201 LeuLeuCySPProProProArgLeuCySPThrAspAsnGlyIleMetIleAlaIleAsnGly 220
DB 601 TTGTGTGTCTCTCCCTCCAGACTATGCACTGATTAATGCAATTAATGCAATGCAATGCT 660
QY 221 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyTyrlleArgTyrglu 240
DB 661 ATTGAAGACTACGTGCTGGCTTGAGCATTTTACATGACATGAAAGGCATCCGCTATGAA 720
QY 241 ProLysCySPProLeuGlyValAspIleSerLysGluValGlyGluAlaSerIleLysVal 260
DB 721 CCAAAATGCTCTTGGAGTAGACATATCAAAAGAAAGTTGGAAGGCTTCCATTAAGTA 780
QY 261 ProGlnLeuLysMetGluIle 267
DB 781 CCACAAATTAATAATGAGATA 801

RESULT 3
BC011904
LOCUS
DEFINITION
Homo sapiens O-sialoglycoprotein endopeptidase-like 1, mRNA (CDNA
clone MGC:20293 IMAGE:4121450), complete cds.
ACCESSION
BC011904
VERSION
BC011904.2 GI:40225818
KEYWORDS
MGC.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 1908)
Strausberg,R.L., Fetschold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shemmen,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Datchenko,L., Marsina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Schaefer,T.E., Brownstein,M.J., Usdin,T.B., Toshlyuk,I.S.,
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullany,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Morley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahey,J., Helton,E., Kettelman,M., Madan,A., Rodriguez,S.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smalls,D.E.,
Schmerch,A., Schein,J.E., Jones,S.J., and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
2 (bases 1 to 1908)
Strausberg,R.
Direct Submission
Submitted (30-JUN-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
On Dec 19, 2003 this sequence version replaced gi:15080281.
REMARK
COMMENT
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: <http://www.nisc.nih.gov/>
Contact: nisc_mgc@hgt.nih.gov
Ahter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
Blakesley,R.W., Bouffard,G.G., Breun,K., Brinkley,C., Brooke,S.,
Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
Hansen,N., Ho,S.-L., Karins,E., Kwong,P., Latic,F., Legaaspi,R.,
Maduro,Q.L., Mastello,C., Masker,B., Mastrian,S.D., McCloskey,J.C.,
McDowell,J., Pearson,R., Stantirlop,S., Thomas,P.J., Touchman,J.W.,
Tsurgeon,C., Vogt,J.L., Walker,M.A., Weherby,K.D., Wiggins,L.,
Young,A., Zhang,L.-H. and Green,E.D.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LMNL at: <http://image.lnl.gov>
Series: IRAL Plate: 28 Row: 1 Column: 22.
Location/Qualifiers
1. 1908
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:20293 IMAGE:4121450"
/tissue_type="Muscle, rhabdomyosarcoma"

misc_feature
/clone_11b="NIH_MGC_17"
/lab_host="DH10B-R"
/note="vector: pOTB7"
1. 1908
/gene="OSGEP1"
/db_xref="LOCUSID:64172"
104. 1348
/gene="OSGEP1"
/codon_start=1
/product="OSGEP1 protein"
/protein_id="AAH1904.1"
/db_xref="GI:15080282"
/db_xref="LOCUSID:64172"
/translation="MLILTKIVGPFKSKRYVEFLRSFNHPGTLFLAKIVLGIET
SCDDPAAAVDETVNLGLAHSQTEVHLTKGTIVPAQAQIPIHHEHAIITRLTNK
VSPFDLSAIIITIGALSLGVGLSFSLGVGLGKPEPIPHHEHAIITRLTNK
VSPFDLSAIIITIGALSLGVGLSFSLGVGLGKPEPIPHHEHAIITRLTNK
MSGRAIETLAKQNRHFDIPPLHAKNCPSFGLOHVTDKITMKKEBGLKQ
QILSSAADIATVQHTMACHVLRTHRALIFCKQDLRPNNAVAVASGVASNFYR
RALILTNATQCTLCPPRLCTDNGIMTAMNGIERLRAGLILHDIIEGIRYEPKPL
GVDSIKVEVEASIKVPLQKMEI"
215. 1111
/gene="OSGEP1"
/note="peptidase M22, Region: Glycoprotease family"
/db_xref="CDD:pfam00814"
ALIGNMENT Scores:
Pred. No.: 1,976-126 Length: 1908
Score: 1385.00 Matches: 267
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0
US-10-649-273-2_COPY_148_414 (1-267) x BC011904 (1-1908)
QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnIleValGluPheProPheLeuVal 20
DB 545 ATGGAGGCTCAGCAGCTTACTTTCAGGTTGACCAATTAAGTAAATTTCTTTTATGTT 604
QY 21 LeuLeuIleSerGlyIleHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
DB 605 CTTTGGATTTCAGAGTCACGTCTGTTGGCATTAGTTCAAGAGATTTCAGATTTTCTG 664
QY 41 LeuLeuGlyIleSerLeuAspIleAlaProGlyAspMetLeuAspIleValAlaArg 60
DB 665 CTTTGGAAAGCTCTTGGACATGACACCGATGACATCTTGACAAAGGTGCAGAGAA 724
QY 61 LeuSerLeuIleLeuHisProGluCysSerThrMetSerGlyGlyValAlaIleGluHis 80
DB 725 CTTTCTTTAATTAACATCCAGAGTCTCCACAGATGAGTGGAGAAAGCCATTGACAT 784
QY 81 LeuAlaIleGlnGlyAsnArgPheHisPheAspIleLeuAspProProLeuHisIleAlaIle 100
DB 785 TTGGCAAAACAGAAATAGATTTTCAATTTGACATCAACCTCCCTTGATCATGCTAA 844
QY 101 AsnCyAspPheSerPheThrGlyLeuGlnHisValThrAspIleIleIleMetIleVal 120
DB 845 AATTGTGATTTTCTTTTACTGAGACTTCAACACGTTACTGATTAATAATATATGAAAAG 904
QY 121 GluIleGlnGluGluGlyIleGluIleGluGlnIleLeuSerSerIleAlaAspIleAla 140
DB 905 GAAAAAGAGAAAGATTAAGAAAGGGCAAAATCTGTCTTCAGCAGCAGCAACATTCGTGC 964
QY 141 ThrValGlnHisThrMetAlaCysHisLeuValIleValArgThrHisArgAlaIleLeuPhe 160
DB 965 ACAATACAGCACAACATGATGATCTTGTGAAAAGAACACATCGGCTATTTCTGTT 1024
QY 161 CysIleGlnArgAspLeuLeuProGlnAsnAsnAlaValIleValAlaSerGlyIleVal 180
DB 1025 TGTAAAGACAGAGAGACTTGAATCAATAATATGACAGTACTGTTGATCTGTGTGTGTC 1084

QY 181 AlAsersanpheyrlleargargalaleuGlulieleuthrAsnAlaIthrGlnCysThr 200
DB 1085 GCAAGTAACCTTCTATTCGCGAGAGCTCGGAAATTTTAAACAAACGCAACAGAGCACT 1144
QY 201 LeuLeuCySPROPORProArgLeuCythrAspAsnglyllemerllealatrAsnly 220
DB 1145 TTGTTGTCCTCTCCACAGCTATGACATGATTAATGCAATTATGATTCAGAGAAATGCT 1204
QY 221 lleGlualrGlueaAGAlaglyleuGlylleuHIsAspIleGlulyleargTyrglu 240
DB 1205 ATTGAAGACTACGTCGCTGGCATTTTACATGACATGAAAGGACATCGCTATGAA 1264
QY 241 ProlysCySPROleuGlyValAspIleSerlysgluValGlyglualaserllelysVal 260
DB 1265 CCAAAATGTCCTCTTGAGTAGACATATCAAAAGAAAGTTGAGAGAGCTTCCATAAAGTA 1324
QY 261 ProGlnLeuLysMetGluIle 267
DB 1325 CCACAAATTAATAATGAGAGATA 1345
RESULT 4
AR428803
LOCUS AR428803 2197 bp DNA linear PAT 18-DEC-2003
DEFINITION Sequence 1 from patent US 6642041.
ACCESSION AR428803
VERSION AR428803.1 GI:40188589
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2197)
AUTHORS Chen,J., Feder,J.N., Nelson,T.C., Krystek,S.R. and Duclos,F.
TITLE Polynucleotides encoding a novel metalloprotease, MP-1
JOURNAL Patent: US 6642041-A 1 04-NOV-2003;
FEATURES
source Location/Qualifiers
1..2197
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Alignment Scores:
Pred. No.: 2,33e-126 Length: 2197
Score: 1385.00 Matches: 267
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0
US-10-649-273-2_COPY_148_414 (1-267) x AR428803 (1-2197)
QY 1 MetGluAlaHIsAlaLeuThrllleargLeuThrAsnlysValGluPheProPheLeuVal 20
DB 672 ATGAGAGGCTCATGCACTTACTATTAGGTGACCAATTAAGTGAATTCCTTTTTCAGTT 731
QY 21 LeuLeuIleSerGlyGlyHIsCysLeuLeuAlaLeuValGlnGlyValaserAspPheLeu 40
DB 732 CTTTGTATTTCGAGGCTGCTGCTGTTGGCATTAGTTCAAGAGACTTTCAGATTTTCTG 791
QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaargarg 60
DB 792 CTTCTTGAAAGGCTTTTGACATGACACAGGTGACATGCTTGACAGAGCTGGCAAGAA 851
QY 61 LeuSerLeuIleLysHIsProGluCysSerThrMetSerGlyLysAlaIleGluHIs 80
DB 852 CTTTCTTAATAATAAACATCCAGAGTGCCTCCACATGAGTGTGGGAAAGCCATAGAACAT 911
QY 81 LeuAlaLysGlnGlyAsnArgPheHIsPheAspIleLysProProLeuHIsHIsAlaLys 100
DB 912 TTGGCAAAACAAAGAAATTAATTCATTTTGACATCAAACTCCCTTGACATCATGTCTAA 971
QY 101 AsnCyAspPheSerPheThrGlyLeuGlnHIsValThrAspLysIlelleMetLysLys 120
DB 972 AATTGTGATTTTCTTTTACTGGAAGCTTCAACAGCTTACTGATTAATAATATGAAAAAG 1031

QY 121 GluLysGlnGluGlylleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAla 140
DB 1032 GAAAAAGAGAAAGGATTCAGAAAGGGCAAACTGCTTTCAGACAGACAACTTCCTGCC 1091
QY 141 ThrValGlnHIsThrMetAlaCysHIsLeuValLysArgThrHIsArgAlaIleLeuPhe 160
DB 1092 ACAGTACAGCACACAATGGCAGTCATCTTGAAAGAAACATCCGGGCTATTCGTTT 1151
QY 161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaserGlyGlyVal 180
DB 1152 TGTAAACAGAGACTTGTACTCAAAATTAATGCAATGCTGTTGATCTGCTGATGTC 1211
QY 181 AlAsersanpheyrlleargargalaleuGlnIleleuthrAsnAlaIthrGlnCysThr 200
DB 1212 GCAAGTAACCTTCTATTCGCGAGAGCTCGGAAATTTTAAACAAACGCAACAGAGCACT 1271
QY 201 LeuLeuCySPROPORProArgLeuCythrAspAsnglyllemerllealatrAsnly 220
DB 1272 TTGTTGTCCTCTCCACAGCTATGACATGATTAATGCAATTATGATTCAGAGAAATGCT 1331
QY 221 lleGlualrGlueaAGAlaglyleuGlylleuHIsAspIleGlulyleargTyrglu 240
DB 1332 ATTGAAGACTACGTCGCTGGCATTTTACATGACATGAAAGGACATCGCTATGAA 1391
QY 241 ProlysCySPROleuGlyValAspIleSerlysgluValGlyglualaserllelysVal 260
DB 1392 CCAAAATGTCCTCTTGAGTAGACATATCAAAAGAAAGTTGAGAGAGCTTCCATAAAGTA 1451
QY 261 ProGlnLeuLysMetGluIle 267
DB 1452 CCACAAATTAATAATGAGAGATA 1472
RESULT 5
AR428808
LOCUS AR428808 1387 bp DNA linear PAT 18-DEC-2003
DEFINITION Sequence 21 from patent US 6642041.
ACCESSION AR428808
VERSION AR428808.1 GI:40188594
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1387)
AUTHORS Chen,J., Feder,J.N., Nelson,T.C., Krystek,S.R. and Duclos,F.
TITLE Polynucleotides encoding a novel metalloprotease, MP-1
JOURNAL Patent: US 6642041-A 21 04-NOV-2003;
FEATURES
source Location/Qualifiers
1..1387
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Alignment Scores:
Pred. No.: 5.43e-124 Length: 1387
Score: 1358.50 Matches: 266
Percent Similarity: 91.10% Conservative: 0
Best Local Similarity: 91.10% Mismatches: 1
Query Match: 98.09% Indels: 25
Gaps: 1
US-10-649-273-2_COPY_148_414 (1-267) x AR428808 (1-1387)
QY 1 MetGluAlaHIsAlaLeuThrllleargLeuThrAsnlysValGluPheProPheLeuVal 20
DB 465 ATGAGAGGCTCATGCACTTACTATTAGTTGACCAATTAAGTGAATTCCTTTTTCAGTT 524
QY 21 LeuLeuIleSerGlyGlyHIsCysLeuLeuAlaLeuValGlnGlyValaserAspPheLeu 40
DB 525 CTTTGTATTTCGAGGCTGCTGCTGTTGGCATTAGTTCAAGAGACTTTCAGATTTTCTG 584
QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaargarg 60

Db 585 CTTCTTGGAAAGTCTTTGGACATAGCACCAGGTGACATGCTTGGACAGGTGGCAAGACA 644

Qy 61 LeuSerleuIlelyHisProGluCySerThrMetSerGlyGlyValaIleGluHis 80

Db 645 CTTTCTTTAATAAATCAATCCAGAGTGTCCACCATGAGTGGGAGAAAGCCATAGAGCAT 704

Qy 81 LeuAlaIlySGInGlyAsnArgPheHisPheAspIlelyProProLeuHisHisAlaIly 100

Db 705 TTGGCCAAACAGAGAAATGATTTTCATTTTGACATCAAACTCCCTTGATCATCTAA 764

Qy 101 AsnCyAspPheSerPheThrGlyLeuGlnHisValThrAspIleIleMetIlyIly 120

Db 765 AATTGTGATTTTCTTTTACTGAGCTTCAACAGCTTACTGATTAATATATAGAAAAAG 824

Qy 121 GluIlySGInGlyIlyle-----GluIlySGInGlyIlyleLeuSerSerAla 126

Db 825 GAAAAAGGAAAGATATTTCTAATAGTAAAGTGAACAGATAATATTCCTGGATTG 884

Qy 127 -----GluIlySGInGlyIlyleLeuSerSerAla 135

Db 885 TGCCTAATAATAGCTGCTCATTTTTCAGAGTATGAAAGGGGCAATCCTGCTTCAGCA 944

Qy 136 AlaAspIleAlaIleThrValGlnHisThrMetAlaCyHisIleValIlyArgThrHis 155

Db 945 GCAGACATTCGCTGCCACGATACAGCACAAATGGCATGTCATTTGTGAAAAAGAACAT 1004

Qy 156 ArgAlaIleLeuPheCyIlySGInArgAspLeuLeuProGlnAsnAsnAlaValIleVal 175

Db 1005 CGGGCTATTCCTGTTTGTGAACAGAGACATTTGTTACTCAAAATATTCAGTACTGGTT 1064

Qy 176 AlaSerGlyIlyValAlaSerAsnPheIlyIleArgArgAlaLeuGlnIleLeuThrAsn 195

Db 1065 GCATCTGGTGGTGTGCGCAAGTAACTTCTATATCCGCAAGCTCTGAAATTTTAAACAAC 1124

Qy 196 AlaThrGlnCySerThrLeuLeuCySerProProAlaGlyLeuCyThrAspAsnGlyIlyIleMet 215

Db 1125 GCACACAGAGTCACTTGTGTGTCTCTCCCAACATGATGCAATGATGCAATGATGCAATGATG 1184

Qy 216 IleAlaIleArgPheGlyIleGluArgLeuArgAlaGlyIlyleGluHisAspIleGlu 235

Db 1185 ATTCATGAGATGATGATTTGAAAGACTACGTGCTGGGCAATTTTACATGACATAGAA 1244

Qy 236 GlyIleArgIlyGlyProIlyCySerProLeuGlyValAspIleSerIlyGlyValGlyIly 255

Db 1245 GGATCCGCTATGACCAAAATGTCCTTGGAGTGAACATATCAAAAAGATTGAGAGA 1304

Qy 256 AlaSerIleIlyValProGlnLeuIlyMetGluIly 267

Db 1305 GCTTCCATAAAAGTACCAATTAATAATGAGAGATA 1340

RESULT 6

LOCUS HSA295148 1387 bp mRNA linear PRI 30-OCT-2000

DEFINITION Homo sapiens mRNA for putative sialoglycoprotease type 2.

ACCESSION AJ295148.1 GI:11071726

VERSION 1

KEYWORDS metalloproteinase; sialoglycoprotease.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 Chen, J.M., Fortunato, M. and Barrett, A.J. Cloning and sequencing of a second human putative sialoglycoprotease homologue unpublished

JOURNAL 2 (bases 1 to 1387)

AUTHORS Chen, J.M.

TITLE Direct Submission

REFERENCE Submitted (27 OCT-2000) Chen J.M., McC Molecular Enzymology Laboratory, The Babraham Institute, Babraham, Cambridge, CB2 4AT, UNITED KINGDOM

FEATURES

Location/Qualifiers

source 1. 1387

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/feature_type="utermus"

24. 1343

/function="metalloproteinase of family M22"

/codon_start=1

/product="putative sialoglycoprotease type 2"

/protein_id="CAC14666.1"

/db_xref="GI:11071727"

/db_xref="COA:Q9H480"

/db_xref="UniProt/TREMBL:Q9H480"

/translation="MLITKTAGVFPKPSKRYEFLSFPNFPETLLFLAKIVLGIETSCDDPAVAVDEGTGVLGEAHSQTEVHLKTGTVIPPAQQLHRENIQRIQVDEALSASGVSPSLDIATTKIPKGLSLGVGSFELQVGLGKKEPIPHHEAVALFRLTNKVFPPFLVILISGHCILALVQVSDFLILGKSLIDPGLMDLVARSLIKIPBCSTMSGKRAIEHLAQGRFHDIPRLHAKRCPSFGLQVHTDKITMKEKEGIFLISKEQINIFGLCKIAHFCRYEKQILSSADIATVQHTMACHLVKKRTHRAILFCKQRDLPPONNAVIVASGVASNFYIRRALIILTNATQCTLLCPRPRLCTDNGIMIAWNGIERLKGILGILHIDBIGIRYRXPCLGVDISKEVGEASIRVPLQKMEI"

ORIGIN

Alignment Scores:

Pred. No.: 5,43e-124 Length: 1387

Score: 1358.50 Matches: 266

Percent Similarity: 91.104 Conservative: 0

Best Local Similarity: 91.104 Mismatches: 1

Query Match: 98.098 Indels: 25

DB: 9 Gaps: 1

US-10-649-273-2_COPY_148_414 (1-267) x HSA295148 (1-1387)

Qy 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnIlyValGluPheProPheLeuVal 20

Db 465 ATGAGAGGCTCAGTACCTTACATTAAGTGTGACCAATTAAGTAAATTTCTTTTAAAGT 524

Qy 21 LeuLeuIleSerGlyGlyHisCySerLeuAlaLeuValGlnIlyValSerAspPheLeu 40

Db 525 CTTTGTGATTTCTGGAGTACTGTCTGTGGCATTTGTTCAAGAGATTTCAGATTTTCTG 584

Qy 41 LeuLeuGlyIlySerIleAspIleAlaProGlyAspMetLeuAspIlyValAlaArgArg 60

Db 585 CTTCTTGGAAAGTCTTTGGACATAGCACCAGGTGACATGCTTGAACAAGTGGCAAGA 644

Qy 61 LeuSerleuIlelyHisProGluCySerThrMetSerGlyGlyValaIleGluHis 80

Db 645 CTTTCTTTAATAAATCAATCCAGAGTGTCCACCATGAGTGGGAAAGCCATAGAGCAT 704

Qy 81 LeuAlaIlySGInGlyAsnArgPheHisPheAspIlelyProProLeuHisHisAlaIly 100

Db 705 TTGGCCAAACAGAGAAATGATTTTCATTTTGACATCAAACTCCCTTGATCATCTAA 764

Qy 101 AsnCyAspPheSerPheThrGlyLeuGlnHisValThrAspIleIleMetIlyIly 120

Db 765 AATTGTGATTTTCTTTTACTGAGCTTCAACAGCTTACTGATTAATATATAGAAAAAG 824

Qy 121 GluIlySGInGlyIlyle-----GluIlySGInGlyIlyleLeuSerSerAla 126

Db 825 GAAAAAGGAAAGATATTTCTAATAGTAAAGTGAACAGATAATATTCCTGGATTG 884

Qy 127 -----GluIlySGInGlyIlyleLeuSerSerAla 135

Db 885 TGCCTAATAATAGCTGCTCATTTTTCAGAGTATGAAAGGGGCAATCCTGCTTCAGCA 944

Qy 136 AlaAspIleAlaIleThrValGlnHisThrMetAlaCyHisIleValIlyArgThrHis 155

Db 945 GCAGACATTCGCTGCCACGATACAGCACAAATGGCATGTCATTTGTGAAAAAGAACAT 1004

Qy 156 ArgAlaIleLeuPheCyIlySGInArgAspLeuLeuProGlnAsnAsnAlaValIleVal 175

Db 1005 CGGGCTATTCCTGTTTGTGAACAGAGACATTTGTTACTCAAAATATTCAGTACTGGTT 1064

QY 176 AlAserGlyValAlaSerAsnPhetYrIleargAlaLeuGluIleLeuThrAsn 195
|
|
|
Db 1065 GCATCTGGTGGTGGCAAGTAATTCATATCCGACAGCTCTGGAATTTTAAACAAC 1124
|
QY 196 AlArthrGlnCysThrIleLeuLeuCysProProAlaGluCysThrAspAsnGlyIleMet 215
|
|
|
Db 1125 GCAACACAGGACCTTGGTGTGTCTCTCCCAACATGACATGATGAATGACATTTG 1184
|
QY 216 IlleAlArpAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlu 235
|
|
|
Db 1185 ATTGCATGGAATGGTATTGAAAGACTACGTGGCTGGCATTTTACATGACATAGAA 1244
|
QY 236 GlyIleArgIYrGluProIYsCysProLeuGlyValAspIleSerIYsGluValGlyGlu 255
|
|
|
Db 1245 GGCAATCGCTATGAAACCAAAATGTCTCTTGGAGTAGACATATCAAAAGAAAGTTGAGAA 1304
|
QY 256 AlAserIleIYsValProGluIleuIYsMetGluIle 267
|
|
|
Db 1305 GCTTCATMAAAGTAAACCAATTTAAATGAGATTA 1340
|
RESULT 7
AX664697 1245 bp DNA linear PAT 22-MAR-2003
LOCUS AX664697
DEFINITION Sequence 6 from Patent WO02074960.
ACCESSION AX664697
VERSION AX664697.1 GI:29164457
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1
AUTHORS Leiby K.R., Kapeller-Libermann, R. and Glucksmann, M.
TITLE 38650, 28472, 5495, 65507, 81588 and 14354 methods and compositions
JOURNAL of human proteins and uses thereof
Milestone Pharmaceuticals, Inc. (US)
FEATURES
source 1. 1245
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Alignment Scores:
Pred. No.: 5.14e-123 Length: 1245
Score: 1348.00 Matches: 260
Percent Similarity: 98.50% Conservative: 3
Best Local Similarity: 97.38% Mismatches: 4
Query Match: 97.33% Indels: 0
Gaps: 0
US-10-649-273-2_COPY_148_414 (1-267) x AX664697 (1-1245)
QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnIYsValGluPheProPheLeuVal 20
|
|
|
Db 442 ATGAGGCTCATGCACCTTACTTATTTAGCTTGACCAATAAAGTGAATTTCTTTTAGTT 501
|
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
|
|
|
Db 502 CTTTGTGATTTCTGGAGGTCACGTGCTGTGGCATTAGTTCAAGGATTTTCAATTTTCTG 561
|
QY 41 LeuLeuGlyIYsSerIleuAspIleAlaProGlyAspMetLeuAspIYsValAlaArgArg 60
|
|
|
Db 562 CTTTGTGAAAGTCTTTTGACATACACACAGGTGACATGCTTGACACAGGTGCAAGAA 621
|
QY 61 LeuSerIleuIleYsHisProGluCysSerThrMetSerGlyGlyIYsAlaIleGluHis 80
|
|
|
Db 622 CTTTCTTTAAATAAACAATCCAGAGTCTCACCAATGAGTGAGGAAAGCCATAGAACT 681
|
QY 81 LeuAlaIYsGlnGlyAsnArgPheHisPheAspIleIYsProProLeuHisHisAlaIYs 100
|
|
|
Db 682 TTGGCAAAACAAGAAATAGATTTCATTTTGAATCAAACTCCCTTGATCATGTCTAAA 741
|

QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspIYsIleIleMetIYsIYs 120
|
|
|
Db 742 AATTGGAATTTTCTTTTACTGCACTTCAACAGCTTACATGATPAAAATATGAAACAG 801
|
QY 121 GlnIYsGluGluGlyIleGluIYsGlyGlnIleLeuSerSerAlaAlaAspIleAlaIle 140
|
|
|
Db 802 AAACAAGAGGAAGGATTTAGAAAGGGCAAACTCTGTCTTCAGCAGCAGCATTTGCTGCC 861
|
QY 141 ThrValGlnHisThrMetAlaCysHisIleuValIYsArgThrHisArgAlaIleLeuPhe 160
|
|
|
Db 862 ACGATGACACACAAATGCAATGTCACTTTTGAAAAGAAACATGGGCTATTCTGTTT 921
|
QY 161 CysIYsGlnArgAspLeuLeuProGlnAsnAlaValIleuValAlaSerGlyIYsVal 180
|
|
|
Db 922 TGTAAAGAGAGAGACTTGTATCTGAAAATAATGCAGTATCGTTGCATCTGGTGGTGC 981
|
QY 181 AlAserAsnPhetYrIleargAlaLeuGluIleLeuThrAsnAlArthrGlnCysThr 200
|
|
|
Db 982 GCAGTAATCTCTATATCCGACAGCTCTGAAATTTTAAACAACCAACACAGTGCACCT 1041
|
QY 201 LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlArpAsnGly 220
|
|
|
Db 1042 TTGTTGTGTCTCTCCCAAGCTACGATGATTAATGCAATTAATGATGCAATGATGCT 1101
|
QY 221 IlleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgIYrGlu 240
|
|
|
Db 1102 ATTGAAAGCTACGTCCTGGCTGGCATTTTACATGACATGAAAGCAATCCGCTATGAA 1161
|
QY 241 ProIYsCysProLeuGlyValAspIleSerIYsGluValGlyGluAlaSerIleIYsVal 260
|
|
|
Db 1162 CCAAAATGTCTCTTGGAGTAGACATATCAAAAGAAAGTTGAGAAAGCTTCATMAAAGTA 1221
|
QY 261 ProGluIleuIYsMetGluIle 267
|
|
|
Db 1222 CCACATTTAAATAATGAGATTA 1242
|
RESULT 8
AX664695 1820 bp DNA linear PAT 22-MAR-2003
LOCUS AX664695
DEFINITION Sequence 4 from Patent WO02074960.
ACCESSION AX664695
VERSION AX664695.1 GI:29164455
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1
AUTHORS Leiby K.R., Kapeller-Libermann, R. and Glucksmann, M.
TITLE 38650, 28472, 5495, 65507, 81588 and 14354 methods and compositions
JOURNAL of human proteins and uses thereof
Milestone Pharmaceuticals, Inc. (US)
FEATURES
source 1. 1820
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
/note="unnamed protein product"
/codon_start=1
/protein_id="CAD80044.1"
/db_xref="GI:29164456"
/translation="MLILKTKAGVFFPKRKRYEFELRSFNFHPTLFLHKIVLGLET
SCDDPAAAVDEGTGVLGEAIIHQTEVNHKTGIVPPAQOLHRENIQRTVQALASAS
GVSPDLVAIATITIKGALSIGVGSFSLQIVGQAKKPIPIHHMEALHLITRLTNK
VRRPPIYLLISGHCLIALVOGVSDFLITGKSIDIDAGMDLDAVARLSLTKHPEST
MSGKRAIEHLAOGKRFHFDIKRPLHAKKCDSPFGLOHTYTDKNENPROEGEIKG
QLLSADIAATVQHTWACHLVKRTIRALIFKQRODLLPQNNAVLVAAGVAVSNFTIR
RALBITVNAVQCTLLCPRPRLCTDNGIMTAMNGIEBRLRGLGIHDI EGIRYPKCPL
GVDISKEVEASIKVPLQKMEI"
ORIGIN

Alignment Scores:

Pred. No.: 8,17e-123 Length: 1820
 Score: 1348.00 Matches: 260
 Percent Similarity: 98.50% Conservative: 3
 Best Local Similarity: 97.38% Mismatches: 4
 Query Match: 97.33% Indels: 0
 DB: 6 Gaps: 0

US-10-649-273-2_COPY_148_414 (1-267) x AX646695 (1-1820)

```

QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLeuValGluPheProPheLeuVal 20
DB 587 ATGAGAGCTCATGACCTTACTATTAGGTTGACCAATTAAGAAATTTCTTTTACTT 646
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
DB 647 CTTTGTATTCTGAGAGCTACTGCTCTGTGGCATTAGTTCAAGAGATTTCAGATTTCG 706
QY 41 LeuLeuGlyIleSerLeuAspIleAlaProGlyAspMetLeuAspIleValAlaArgArg 60
DB 707 CTTCTTGAAAGTCTTTGACATGACACAGGTGACATGCTTGACAAAGTGGCAAGAAGA 766
QY 61 LeuSerLeuIleIleHisProGluCysSerThrMetSerGlyGlyValAlaIleGluHis 80
DB 767 CTTTCTTTAATTAACATCCAGAGTCTCCACCAATGAGTGGGAAAGCCATAGAACAT 826
QY 81 LeuAlaIleGlnGlyAsnArgPheHisPheAspIleIleValProProLeuHisAlaIle 100
DB 827 TTGGCCAAACAGAAATAGATTTCATTTCATTTCACCAAACTCCCTTCATCATGCTTAA 886
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValIleThrAspIleIleIleMetIle 120
DB 887 AATTGTATTCTTTTACTTGTGACCTTCAACGCTTACGTAATTAATAAGAAACAGG 946
QY 121 GluIleGlnGluGlyIleGluIleGlyGlnIleLeuSerSerAlaAlaAspIleAlaAla 140
DB 947 AAACAAAGAGAAAGTATTGAAAGAGGCAAAATCTGCTTCAGCAGCAGCATTCCTGCC 1006
QY 141 ThrValGlnHisThrMetAlaCysHisIleValIleValIleArgThrHisAlaIleLeuPhe 160
DB 1007 ACAGTACAGCACACAAATGCGATGTCATCTTGTGAAAACACATCGGCGTATTCTGTTT 1066
QY 161 CysIleGlnArgAspLeuLeuProGlnAsnAsnAlaValIleLeuAlaIleSerGlyVal 180
DB 1067 TGTAGAGAGAGAGCTTGTACTCTCAAAATTAATGAGTACTGCTGTCATCTGCTGCTGC 1126
QY 181 AlaSerAsnPheIleIleArgArgAlaLeuGluIleLeuThrAsnAlaIleThrGlnCysThr 200
DB 1127 GCAAGTAACTTCTATATCCGAGAGCTCTGGAATTTTAACAAACGCAACACAGTCACT 1186
QY 201 LeuLeuCysProPheProIleArgLeuCysThrAspAsnGlyIleMetIleAlaIlePheAsnGly 220
DB 1187 TTGTGTGTCTCTCTCCAGACTATGCACTATATATGCACTATATGCACTATGGAATGCT 1246
QY 221 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlnGlyIleIleArgThrGlu 240
DB 1247 ATTGAAAGACTACTGCTGCTGCTTGGCATTTTACATGACATGAAAGGCAATCCGCTATGAA 1306
QY 241 ProIleCysProLeuGlyValAspIleSerIleGlnValGlyIleAlaIleSerIleIleVal 260
DB 1307 CCAAAATGCTCTCTTGGAGTACATATCAAAAGAGTTGAGAGAGCTTCCATTAAGTA 1366
QY 261 ProGlnLeuIleValMetGluIle 267
DB 1367 CCACAAATTAATAATGAGATTA 1387

```

KEYWORDS

SOURCE

MGC.
 Mus musculus (house mouse)
 Mus musculus

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 1844)

REFERENCE

AUTHORS

Strausberg, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
 Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
 Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Heich, F.,
 Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
 Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
 Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
 Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
 Abramson, R.D., Mullany, S.J., Bosak, S.A., McGwan, P.J.,
 McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.W.,
 Morley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hultk, S.W.,
 Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
 Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S.,
 Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
 Bouffard, G.G., Blakeley, R.W., Touchman, D.W., Green, E.D.,
 Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
 Butlerfield, Y.S., Krzywinski, M.I., Skalska, U., Smalios, D.E.,
 Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.
 Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 22368257
 12477932
 2 (bases 1 to 1844)
 Strausberg, R.
 Direct Submission
 Submitted (15-SEP-2003) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Offices, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA

REMARK

NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgabs-remail.nih.gov

COMMENT

Tissue Procurement: Gilbert Smith, Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)
 DNA Sequencing by: National Institutes of Health Intramural
 Sequencing Center (NISC),
 Gaithersburg, Maryland;
 Web site: <http://www.nisc.nih.gov/>
 Contact: nisc_mgc@ngri.nih.gov
 Akherley, N., Ayle, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
 Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
 Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
 Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Latic, P., Legaspi, R.,
 Maduro, Q.L., Masiello, C., Maskell, B., Mastrian, S.D., McCloskey, J.C.,
 McDowell, J., Pearson, R., Scantirip, S., Thomas, P.J., Touchman, J.W.,
 Tsugeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
 Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/BLIML at: <http://image.llnl.gov>
 Series: IRAX Matrix: 123 Row: d Column: 22
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 21312463.
 Location/Qualifiers

FEATURES

source

1..1844
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="MGC:67870 IMAGE:5012054"
 /issue_type="Mammalian tumor metastatic to lung, tumor
 arose spontaneously from a senescent normal mammary
 (clonal) outgrowth infected with the virus MMTV."
 /clone_lib="NCI CGAP_Lu29"
 /lab_host="DH10B"

CDS

/note="Vector: PCMV-SPORT6"
222..1466
/codon_start=1
/product="unknown (protein for MGC:67870)"
/protein_id="AA58172.1"
/db_xref="GI:34849664"
/translating="MMLRRYAGAI PKPKSKYVGLRR FSVHPRTLSCHKLVIGIET
SCDPTGAIVDEGTGVALSOTQVHFKGIVP PAQOOLHRENIORIVETLSAS
RTPSDLSAIVTTIKPGLALSLGVLSLQVLRFKKPIPIIHMEHALITRLTK
VEPPEVLILSGCHLALVQGVSDPILGLDIAQDMLDKVARRLSLIKHPCST
MSGKALIEHLAKDKRFRFTINPWNQKNCDFSTGLOHTTKITKEKEBGEK
QIISADIAAIVQATRAHLAKRTHRALIFCKOKNLISPAVAIVVGGVANSIYIR
KALRIVANATQCTLCPPRLCTDNGIMIANNGIERLRAGLGVLDVEDIRVEPCPL
GIDISREVAEAIKIVRLKML"
333..1397
/note="ORF1; Region: Metal-dependent proteases with
possible chaperone activity [Posttranslational
modification, protein turnover, chaperones]"
/db_xref="CDD:COG0533"

misc_feature

333..1397
/note="ORF1; Region: Metal-dependent proteases with
possible chaperone activity [Posttranslational
modification, protein turnover, chaperones]"
/db_xref="CDD:COG0533"

ALIGNMENT Scores:

Pred. No.:	2.76e-111	Length:	1844
Score:	1231.00	Matches:	233
Percent Similarity:	92.88%	Conservative:	15
Best Local Similarity:	87.27%	Mismatches:	19
Query Match:	88.88%	Indels:	0
DB:	10	Gaps:	0

US-10-649-273-2_COPY_148_414 (1-267) x BC058172 (1-1844)

Qy 1 MetGUAIAHISALAEUThrllearyleuThrasuylvaIGluPheProPheleuVal 20
Db 663 ATGAGGCTCAGCAGCTACCTATTTAGGCTCACCAATAAGTAATTCCTTTCTAGTT 722
Qy 21 LeuLeuIleSerGIyGIYHISCYsteLeuLeuAlaLeuValGInGIYValSerAspPheLeu 40
Db 723 CTTTGTATTTCTGGGGTCACTGCTGTGGCATTTAGTCCAAAGTGTTCCGATTTCCG 782
Qy 41 LeuLeuGIYLySerLeuAspIleAlaProGIYAspMetLeuAspLyValAlaArgATG 60
Db 783 CTCTCTTGAAGTCTTTGGACATACGCGACGACGACATGCTTGACAAAGTGCGCAAGMA 842
Qy 61 LeuSerLeuIleLyHisProGIYucySerThrMetSerGIYLyValAlaIleGIYHis 80
Db 843 CTTTCTTAAATCAACATCCAGAAATGTTCTACAAATGAGTGAGGAAAGCTATAGAACT 902
Qy 81 LeuAlaLySGInGIYAsnArgPheHisPheAspIleLyProProLeuHisIleAlaLyS 100
Db 903 TTGGCCAAAGACGAAATAGATTCCATTTACTATCAATCCATATGCAAGATGCTAAG 962
Qy 101 AsnCyAspPheSerPheThIGLyLeuGInHisValThraspLyIleIleMetLyS 120
Db 963 AATTGCGATTTCTTTACGGGACTTCAACATATTACTGATAGCTATTAACACACAG 1022
Qy 121 GIYLySGInGIYLyIleGIYLySGInIleLeuSerSerAlaAlaAspIleAla 140
Db 1023 GAAAAAAGAAAGGACATTCGAAAGGGCAATCTGTCTATCGACGCAACATGTGCTG 1082
Qy 141 ThrValGIInHisThMetAlaCyHisIleuValLySArgThHisArgAlaIleLeuPhe 160
Db 1083 GCGGTACGATGCAACAGCGTGCACCTTGCMAAAAGAACACATGCTATTCGTGTT 1142
Qy 161 CysLySGInArgAspLeuLeuProGIInAsnAlaValIleuValAlaSerGIYLyVal 180
Db 1143 TGCAGACGAGAAATTTGCTATCTCAGCTAACGAGTATTAAGTTGTTATCTGAGAGTGT 1202
Qy 181 AlaSerAsnPheTyTlleArgArgAlaLeuGIYIleLeuThrasnAlaThrgInCySThr 200
Db 1203 GCAAGTAACTTGATCATCCGAAAGACATTCGAAATTCGCAAAATGCCAAGCTGCAAG 1262
Qy 201 LeuLeuCyProProArgArgLeuCyThrAspAsnGIYIleMetIleAlaTPanGIY 220

Db 1263 TTGCTGTGTCCTCCCTCCAAAGACTGTCAGTCAATGCGATCATGATTCAGTGAATGCA 1322
Qy 221 IlegIuArgLeuArgAlaGIYLeuGIYIleLeuHisAspIleGIYIleArgTyrgIn 240
Db 1323 ATTTGAAAGATTACGTGCTGCTGGCGCTTTTACATGATGATGAGACATCCGATATGAA 1382
Qy 241 ProLyCyProLeuGIYValAspIleSerLySGInValGIYGIYValAlaSerIleLyVal 260
Db 1383 CCAAAATGCTCTTGGAATATGACATATCCAGAGAAATTCGAGAAAGCTCCATAAAAATGA 1442
Qy 261 ProGIInLeuLyMetGIYIle 267
Db 1443 CCGCATTTAAAAATGCGACTT 1463

RESULT 10
BC038910
LOCUS
DEFINITION
MUS musculus O-sialoglycoprotein endopeptidase-like 1, mRNA (cDNA
clone IMAGE:5053559), partial cds.
ACCESSION
BC038910
VERSION
BC038910.1 GI:24433548
KEYWORDS
MUS musculus (house mouse)
SOURCE
MUS musculus
ORGANISM
MUS musculus (house mouse)
REFERENCE
AUTHORS
Strausberg,R.L., Peingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shemmen,C.M., Schler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Scheffer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,D., Hsieh,F.,
Diachenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Donald,M.F., Casavert,T.L.,
Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
Carninci,P., Prange,C., Rana,S.S., Loquellano,N.A., Peters,G.J.,
Adams,R.D., Mullaly,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahey,J., Helton,B., Kettman,M., Madan,A., Rodrigues,S.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Boutfard,G.G., Blakeley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smallus,D.E.,
Scherer,A., Schein,J.B., Jones,S.J. and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
JOURNAL
MEDLINE
PUBMED
22388257
12477932
TITLE
2 (bases 1 to 1017)
AUTHORS
Strausberg,R.
DIRECT SUBMISSION
Submitted (25-OCT-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
infocgsc.bc.ca
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
Susana Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
Leticia Hsiao, Martin Krzywinski, Keta Kutsche, Oliver Lee, Soo
Sen Lee, Victor Ling, Carrie Mahewson, Candace McEay, Steven
Ness, Pawan Pandoh, Anna-Liisa Prabhun, Parvaneh Saadei, Jacqueline
Schein, Duane Smallus, Michael Smith, Lorraine Spence, Jeff Stott,
Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,

George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/IMLML at: <http://image.llnl.gov>
Series: IRAK Plate: 86 Row: f Column: 12.

FEATURES

Source

Location/Qualifiers
1..1017
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5053559"
/issue_type="Liver, normal. 5 month old male mouse."
/clone_id="NCI CGAP_L19"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"

ORIGIN

Alignment Scores:

Pred. No.:	6,53e-111	Length:	1017
Score:	1224.00	Matches:	233
Percent Similarity:	92.51%	Conservative:	14
Best Local Similarity:	87.27%	Mismatches:	20
Query Match:	88.38%	Indels:	0
DB:	10	Gaps:	0

US-10-649-273-2_COPY_148_414 (1-267) x BC038910 (1-1017)

QY 1 MetGUAAlAHlAAlaUeUThrIlleArGleUThrAsnLySValGluPheProPheUVal 20
DB 26 ATGAGAGCTCAAGCAGCTACTATTAGGCTCAACATAAGATGAAATTCCTTTTAACT 85
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaUeUValGlnGlyValSerAspPheU 40
DB 86 CTTTGGATTCTGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 145
QY 41 LeuLeuGlyLySerLeuAspIleAlaProGlyAspMetLeuAspLyValAlaArgArg 60
DB 146 CTCCTTGGAAGCTTTTGACATAGCACAGGCGACATGCTTGACAAAGTGGCAAGAGA 205
QY 61 LeuSerLeuIleLySHisProGlyCysSerThrMetSerGlyGlyValAlaIleGluHis 80
DB 206 CTTTCTTAATCAAAACATCCAGAAATGTTCTCAATGATGGGGCAAAAGCTATAGAACG 265
QY 81 LeuAlaLySGlnGlyAsnArgPheHisPheAspIleUySProProLeuHisAlaLyS 100
DB 266 TTGGCCAAAGACGGAATATGATTCATTCTAATCAATCCATATGACAGATGCTAG 325
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLySIlleUeUcLyS 120
DB 326 AATTGCAATTTTCTTTCACGGGACTTCAACATATTGATGATTAACCTAATTAACACACAG 385
QY 121 GluLySGlnGlyIleGluLySGlyGlnIleLeuSerSerAlaAlaAspIleAla 140
DB 386 GAAAGAAAGAGGACATGAGAGAGGCAAAATCCTGATCAGTCAGACATTCCTCT 445
QY 141 ThrValGlnHisThrMetAlaCysHisLeuValLySArgThrHisArgAlaIleUeU 160
DB 446 GCGGTACAGATGCAACAGCGCTGCACCTTCGCAAAACACACACGCGCTATTCCTTT 505
QY 161 CysLySGlnArgAspLeuUeUProGlnAsnAsnAlaValLeuValAlaSerGlyGlyVal 180
DB 506 TGCAGAGAAATAATTTGCTCTCTCCAGCTAACGAGATTAAGTGTATTCGAGAGCTTT 565
QY 181 AlAserAsnPheTyrlleArgArgAlaUeUValIleUeUThrAsnAlaThrGlnCysThr 200
DB 566 GCAAGTAATCTTGATCCGAAAGACATGGAATATTCGCAAAATGCAACGACAGTCACG 625
QY 201 LeuLeuCysProProArgLeuCysThrAspAsnGlyIleMetIleAlaITrapAsnGly 220
DB 626 TTGTGTGTCCACCTCCAAAGACTGTGCACTGACATGCAATGATGATGCAATGAAATGGA 685
QY 221 IlleGluArgLeuArgAlaGlyLeuGlyIleUeUHisAspIleGluGlyIleArgTyrlu 240

DB 686 ATTGAAGATTACGCTGCTGGCTGCTTTTAAATGATGTAGAGACATCCGATATGAA 745

QY 241 ProLySCysProLeuGlyValAspIleSerLyGlyValGlyGlnAlaSerIleLySVal 260
DB 746 CCAAAATGCTCTTTCGAGTAGACATATCCAGAGAAAGTTCCAGAAAGCTGCCATAAAGTA 805

QY 261 ProGlnLeuLySMetGluIle 267
DB 806 CCGCATTAATAATGACACTT 826

RESULT 11

AX713716
LOCUS AX713716
DEFINITION Sequence 400 from Patent EP1293569.
ACCESSION AX713716
VERSION AX713716.1 GI:29888642
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 Isogai, T., Sugiyama, T., Otsuki, T., Wakamatsu, A., Sato, H., Ishii, S., Yamamoto, J. I., Isono, Y., Hio, Y., Otsuka, K., Nagai, K., Irie, R., Tamechika, I., Seki, N., Yoshikawa, T., Otsuka, M., Nagahari, K. and Masuko, Y.

TITLE

Full-length cDNAs
Patent: EP 1293569-A 400 19-MAR-2003;
Helix Research Institute (JP) ; Research Association for
Biotechnology (JP)

FEATURES

source 1..2208
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Alignment Scores:

Pred. No.:	1.57e-108	Length:	2208
Score:	1204.00 <td>Matches:</td> <td>239</td>	Matches:	239
Percent Similarity:	89.51% <td>Conservative:</td> <td>0</td>	Conservative:	0
Best Local Similarity:	89.51% <td>Mismatches:</td> <td>4</td>	Mismatches:	4
Query Match:	86.93% <td>Indels:</td> <td>24</td>	Indels:	24
DB:	6 <td>Gaps:</td> <td>1</td>	Gaps:	1

US-10-649-273-2_COPY_148_414 (1-267) x AX713716 (1-2208)

QY 1 MetGUAAlAHlAAlaUeUThrIlleArGleUThrAsnLySValGluPheProPheUVal 20
DB 785 ATGAGAGCTCAAGCAGCTACTATTAGGCTCAACATAAGATGAAATTCCTTTTAACT 844
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaUeUValGlnGlyValSerAspPheU 40
DB 845 CTTTGGATTCTGGAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 904
QY 41 LeuLeuGlyLySerLeuAspIleAlaProGlyAspMetLeuAspLyValAlaArgArg 60
DB 905 CTTTCTTAATCAAAACATCCAGAAATGTTCTCAATGATGGGGCAAAAGCTATAGAACG 964
QY 61 LeuSerLeuIleLySHisProGlyCysSerThrMetSerGlyGlyValAlaIleGluHis 80
DB 965 CTTTCTTAATCAAAACATCCAGAAATGTTCTCAATGATGGGGCAAAAGCTATAGAACG 1024
QY 81 LeuAlaLySGlnGlyAsnArgPheHisPheAspIleUySProProLeuHisAlaLyS 100
DB 1025 TTGGCCAAAGACGGAATATGATTCATTCTGACATCAAACTCCCTTGATCATGCTAA 1084
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLySIlleUeUcLyS 120
DB 1085 AATTGGAATTTCTTTTACTGCACTTCAACAGTATGATTAATAATGAAGAAAG 1144
QY 121 GluLySGlnGlyIleGluLySGlyGlnIleLeuSerSerAlaAlaAspIleAla 140

[illegible]

JOURNAL	PUBMED	REFERENCE	AUTHORS	TITLE
Complete sequencing and characterization of 21,243 full-length human cDNAs	Nat. Genet. 36 (1), 40-45 (2004)	14702039	2	Nishii, T., Nakagawa, S., Senoh, A., Mizuguchi, H., Inagaki, H., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Mishi-kawa, T., Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Takuchi, H., Kanda, K., Wagatsuma, M., Murakawa, K., Kanehori, K., Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahara, K., Masuko, Y., Nagai, K. and Isogai, T.
Unpublished				NEDO human cDNA sequencing project
3 (bases 1 to 2208)				Isogai, T., Otsuki, T. and Sugiyama, T.
Direct Submission				Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute, Genomics Laboratory, 1532-3 Yano, Kisarazu, Chiba 292-0812, Japan (E-mail:genomic@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology (RAB); cDNA library construction; Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing; RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: RAB and HRI.				
FEATURES				
source				location/Qualifiers
				1..2208
				/organism="Homo sapiens"
				/mol_type="mRNA"
				/db_xref="taxon:9606"
				/clone="FEBRA2004592"
				/tissue_type="brain"
				/clone_lib="FEBRA2"
				/dev_stage="fetus"
				/note="Cloning vector: pME18SFL3"
				344..1438
				/code="unnamed protein product"
				/codon_start=1
				/protein_id="BAB70923.1"
				/db_xref="GI:16550167"
				/translation="MLILTKTACVFFPSPSKRYEERLRSNFPHDGLFLKILVIGITSCDDTAADVDETNVGEAIIHSQTEVHLTGIVPPAAQOLHRENIQRIQVEXLSASGVSDSLAIAITTIKPGALSLGWSLFSFLQVGLKKPPIPHHMAHALTLRLTKRVERPILVILISGGCLALVOGSDPLILGKSLDIAGMDLKVARELPLIKHECSTMSGKALIEHLAKGNRPHPIKPHHAKNCDPFGLOHYTDKI IKKKEEGEIEKG OIISASADIAIVQHTNACHLVKRTTHAILFCCKRDILLPNNALVVASGVASNFCIRPAELITNATQCTLLCPPLRLCTDNGIMIA"
ORIGIN				
Alignment Scores:				
Pred. No.:	1.57e-108	Length:	2208	
Score:	1204.00	Matches:	239	
Percent Similarity:	89.51%	Conservative:	0	
Best local Similarity:	86.93%	Mismatches:	4	
Query Match:	86.93%	Indels:	24	
DB:	9	Gaps:	1	
US-10-649-273-2_COPY_148_414 (1-267) x AK055441 (1-2208)				
QY	1	MetGuaIaHiaIaLeuThrTleArgLeuThraNlysValIguIpeProPheLeuVal	20	
DB	785	ATGAGGCTCATGCATCTACTATTAGTGTGACCATATAAGTAGAATTCCCTTTTATGTT	844	
QY	21	LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGingIyValSerAspPheLeu	40	
DB	845	CTTTTGATTTCTGAGAGGTCACGTCTGCTGTTGGCAATTAGTTCAAAGAGTTTCAGATTTCGTG	904	
QY	41	LeuLeuGlyIySerSerLeuAspIleAlaIProGlyAspMetLeuAspIyValAlaIArgArg	60	
DB	905	CTTCTGGAAAGCTCTTGGACATGACCCAGAGTACATGCTTGAACAGAGTGGCAAGAA	964	

```

QY      61 LeuSerleuIlelyshisProGluCyseSerThmetserGlylyValaIleGluHis 80
DB      965 CTTCCCTTAATAAACAATCCAGAGTGTCTCCACCATGATGTGGGAAAGCATAGAACAT 1024
QY      81 LeuAlaIleGlnGlnAaenArgPheHisPheAspIleIysPProPolenHisHilaIaIys 100
DB      1025 TTGGCCAAACAGAGAAATAGATTTTCATTTCATTCATCAATCAAACTCCCTTCATCATGCTTAA 1084
QY      101 AenCyAspPheSerPheThrGlyLeuGlnHisValThrAspIleIleIleMetIysIys 120
DB      1085 AATTGTGATTTCTCTTTACTTGACCTTCAACACCTTACTGTATAATATATGAAAAAG 1144
QY      121 GluIleGlnIleGlyIleGlyIleGlyGlnIleLeuSerSerAlaIleAspIleAlaIa 140
DB      1145 GAAAAAGAGAGAGATTTGAGAAAGGGCAAACTCTGTCTTCAGACGACGACATTCCTGCC 1204
QY      141 ThrValGlnHisThrMetAlaCyseHisLeuValIysArgThrHisIleArgAlaIleIleuPhe 160
DB      1205 ACGATACAGACACAAATGGATGATTCATCTGTGAAAAAGAACACATCGGGCTATTCCTGTT 1264
QY      161 CyLeuGlnArgAspLeuLeuProGlnAspAsnAlaValIleValAlaSerGlyGlyVal 180
DB      1265 TGTAAAGACAGAGACTTGTTCCTCAAAATATATGCACTACTGCTGCTGCTGCTGCTGCTC 1324
QY      181 AlaSerAspPheTyrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThr 200
DB      1325 GCAAGTAACTCTTCGTATCCGACAGAGCTCTGGAAATTTTAAACAAACGACACAGTGCACCT 1384
QY      201 LeuLeuCyProProProArgLeuCyseThrAspGlnGlyIleMetIleAlaITTPanGly 220
DB      1385 TTGTTGTGTCCTCTCCACATCTGATGCACTGATGATGATGATGATGATGATGATGATGATGCA 1435
QY      221 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlyIleArgTyrGlu 240
DB      1435 ----- 1435
QY      241 ProIleCyseProLeuGlyValaIleSerIleSerIleGlyIleValaIleSerIleIleVal 260
DB      1436 ---TGATGTCCTCTTGGAGTATGACATATCAAAAGAAAGTTGGAGAGAGCTTCATTAAGTA 1492
QY      261 ProGlnLeuIleYMetGluIle 267
DB      1493 CCACAAATTAATAATGAGATA 1513

RESULT 13
BC078974      1546 bp      mRNA      linear      ROD 03-AUG-2004
LOCUS      Rattus norvegicus cDNA clone IMAGE:7111906, partial cds.
DEFINITION      BC078974
ACCESSION      BC078974
VERSION      BC078974.1 GI:50926879
KEYWORDS
SOURCE      Rattus norvegicus (Norway rat)
ORGANISM      Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 1546)
Strausberg,R.L., Feingold,B.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Schetz,T.B., Brownstein,M.J., Usdin,T.B., Tothiyilki,S.,
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Muliyil,S.J., Bosak,S.A., McKean,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Morley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Vallio,J., Muzny,D.M., Sodergren,K.J., Lu,X., Gibbs,R.A.,
Vahvey,J., Helton,E., Kettman,M., Madan,A., Rodriguez,S.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,

```

JOURNAL
 PUBLISHED
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 2 (bases 1 to 1546)
 Director MGC Project.
 Direct Submission
 Submitted (02-AUG-2004) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: Howard Jacobs
 cDNA Library Preparation: Express Genomics
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome
 Center, Stanford University School of Medicine, Stanford, CA 94305
 Web site: <http://www-shgc.stanford.edu>
 Contact: (Dickson, Mark) mcd@paxil.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
 R. M.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
 Series: IRK Plate: 182 Row: F Column: 6
 This clone was selected for full length sequencing because it
 passed the following selection criteria: Hexamer frequency ORF
 analysis, Similarity but not identity to protein.

FEATURES

source

```

1..1546
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="IMAGE:7111906"
/tissue_type="testis, rat (Brown Norway)"
/clone_lib="NIH MGC_237"
/lab_host="DH10B"
/note="Vector: pExpress1"
144..>1546
/codon_start=1
/product="unknown (protein for IMAGE:7111906)"
/protein_id="AAH78974.1"
/db_xref="GI:50926880"
/translation="MLMLSKTGAALPRPRNSVNRGFRIRFNVQPRALFPHKLVLGIET
SCDDPAAVDEBTGNVLGELALHSOTEVHLKGTGIYVPAQQLHRENIQRIVEALSAS
GVSPSDSLAAITTKDGLASIGVGLSPVOLVNOEKPTPIHMEARLITRLTHK
VGFPPLVILISGHLALVOSVSPFLILGKSLDIAKQMDLKVARRLSLIGPEST
MSGKAIHLAAEGNRFHTINPMDANCDPSFTGLQHVTDKLTTHKEKEGLK
QILSGADIAAVALQATLAKTRHAILFCQOKRLSPAAVALVSGVSNLYIR
RALEIVANAOTCTLCPPRLCTDNGIMAMNGIRLRLGAILHDVEIDIREPKAGE
IVSMRLVLTALTBEDSVSSTHTVAHSPINSSRGANIGTSCSGTQYVYMETVRHTL
NINLKKSKKKKKKKKK"

```

CDS

ORIGIN

Alignment Scores:

```

Pred. No.:      4.7e-100      Length:      1546
Score:          1116.00      Matches:      214
Percent Similarity: 92.56%      Conservative: 10
Best Local Similarity: 88.43%      Mismatches: 18
Query Match:      80.58%      Indels:      0
DB:              10      Gaps:      0

```

US-10-649-273-2_COPY_148_414 (1-267) x BC078974 (1-1546)

```

QY      1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnIleValGluuPheProPheLeuVal 20
DB      585 ATGGAGGCTCAGCGCCTGACTATAGGCTGACCCACAAAGTCGAGATTTCCTTTTAGTT 644

```


QY 21 LeuLeu11eserGlyGlyYhiCysLeuLeu1aleuVal1nglyVal1serAspPheLeu 40
DB 645 CTTTGTGATTTCTGAGAGCCACGCTGTTGGCGTAGTTCAAGAGTTTCAAGATTTTCG 704
QY 41 LeuLeuGly1ysSerLeuAsp11eal1ProGlyAspMetLeuAsp1ysVal1AlaArgArg 60
DB 705 CTCTCGGGAAGTCCCTGAGACATAGCGCAGCGCAGCATGCTTGACAGAGTGGCAAGAA 764
QY 61 LeuSerLeu11elySh1eProGly1uCySerThMetSerGlyGlyVal1leGluHis 80
DB 765 CTTTCTTTAATCAAAACATCCAGAAATGTTCTACAAATGATGGGGAAGAACTTAAGACT 824
QY 81 LeuAl1alysG1ng1yAsnArgPheH1sPheAsp11elysProProLeuH1sh1sAl1alys 100
DB 825 TTGGCCAAAGAAAGAAATAGATTCATCTTACTATCATATCCACCATGAGAAATGCTAAG 884
QY 101 AsnCysAspPheSerPheThrGly1eLeuGlnH1sVal1ThraAsp1ys11e11eMet1ys1ys 120
DB 885 AACTGTATATTTTCTTTTACGCGACTTCAACATGTCACCGATMACCTAATAACACACAAAG 944
QY 121 Glu1ysG1ng1yGly11eGly1uysGly1n1leuSerSerAl1alAsp11eAl1a1a 140
DB 945 GAAAAAGAAAGAGCATTTGAGAGGGGCAATCTCTCATTCAGCCGACATTTGCTGCT 1004
QY 141 ThrVal1GlnH1sThrMetAl1aCySh1sLeuVal1ysArgThrH1sArgAl1al1eLeuPhe 160
DB 1005 GCGGTGACAGCAGCAGCAACGCGTGCACCTTGCGAAAGAAACATCGTCTATTCGTTT 1064
QY 161 Cys1ysG1n1ArgAspLeuLeuProGlnAsnAsnAl1aVal1LeuVal1AlaSerGlyGlyVal1 180
DB 1065 TGCACGAGAAATTTGCTATCTTCACGATTAACGAGATTAATGTTGTGCTGAGAGTGT 1124
QY 181 AlaSerAsnPheThr11eArgArgAl1eLeuGly11eLeuThraSnAl1aThrGlnCysThr 200
DB 1125 GCAAGTAACTTGTATCATCCAGAACATTTGAAATTTAGCAAAATGCAACACAAATGCACT 1184
QY 201 LeuLeuCySProProProArgLeuCySThraAspAsnGly11eMet11eAl1a1ArgAsnGly 220
DB 1185 TTGTGTGTCCCTCCGAGACTGTGCACTGACATGATGATGATGATGATGATGATGATGATGAT 1244
QY 221 11eGlu1ysG1ng1yGly11eGly1uysGly11eLeuH1sAsp11eGly11eArgTyrg1u 240
DB 1245 ATTGAAGATTAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1304
QY 241 Pro1ys 242
DB 1305 CCAAG 1310

RESULT 14
BC051211 1109 bp mRNA linear ROD 15-APR-2003
LOCUS Mus musculus, clone IMAGE:1327545, mRNA.
ACCESSION BC051211
VERSION BC051211.1 GI:29881634
KEYWORDS
SOURCE Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1109)
Strausberg, R.
Direct Submission
Submitted (14-APR-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIR-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabbs-remail.nih.gov
Tissue Procurement: Marcello Bento Soares, Ph.D.
CDNA Library Preparation: M. Bento Soares, University of Iowa
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)
DNA Sequencing by: Institute for Systems Biology

<http://www.systemsbio.org>
contact: amadansystemsbiology.org
Anup Madan, Jessica Fahy, Erin Halton, Mark Kettelman, Anuradha
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/BLN at: <http://image.llnl.gov>
Series: IRAX Plate: 113 Row: b Column: 1
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis.
FEATURES
source location/Qualifiers
1..1109
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:1327545"
/tissue_type="Thymus gland, mouse"
/clone_1lb="Soares_thymus_2NDMT"
/lab_host="DH10B"
/note="Vector: pT73D-Pac"
ORIGIN
Alignment Scores:
Pred. No.: 6.23e-84 Length: 1109
Score: 950.50 Matches: 190
Percent Similarity: 79.10% Conservative: 22
Best Local Similarity: 70.90% Mismatches: 35
Query Match: 68.63% Indels: 21
DB: 10 Gaps: 4
US-10-649-273-2_COPY_148_414 (1-267) x BC051211 (1-1109)
QY 18 PheLeuVal1leuLeu11eser-----GlyGlyH1sCysLeu----- 29
DB 82 TTTTGTGATTTCTGATTTGCGTATTTGCTTGCATTAATTTTGAACATCCCTATATCTCT 141
QY 30 -----LeuAl1eLeuVal1nglyVal1serAspPheLeuLeuGly1ysSerLeuAsp 47
DB 142 TCTGTTTAAACTGACTTGTGTCATATGCGGTTATCTTGTCTTAAT-----AAT 192
QY 48 11eal1eProGlyAspMetLeuAsp1ys-----Val1AlaArg 59
DB 193 TTGCAGCCAAATATATATTTATGAAAAAATGTAATGCTTTTAAATAGTGGCAGAG 252
QY 60 ArgLeuSerLeu11elySh1sProGly1uCySerThMetSerGlyGlyVal1leGlu 79
DB 253 AGACTTTCTTTATCAAAACATCCAGAAATGTTCTACATGATGCTGCGTGAAGAAAGCTATAGAA 312
QY 80 H1sLeuAl1alysG1ng1yAsnArgPheH1sPheAsp11elysProProLeuH1sh1sAl1alys 99
DB 313 CAGTTGGCCAAAGACGAAATAGATTCATTTACTATCATTCACCTATGCGAATGCT 372
QY 100 1ysAsnCysAspPheSerPheThrGly1eLeuGlnH1sVal1ThraAsp1ys11e11eMet1ys 119
DB 373 AAGAAATGGCATTTTCTTCAAGGACCTTCAACATATTAATGATAAGCAATTAACACAC 432
QY 120 1ysGly1uysG1ng1yGly11eGly1uysGly11eLeuSerSerAl1a1aAsp11eAl1a 139
DB 433 AAGGAAAAAGAAAGAGCATTTGAGAGGGGCAAAATCTTCATCACCTGACAGCATTTGCT 492
QY 140 Al1aThraVal1GlnH1sThrMetAl1aCySh1sLeuVal1ysArgThrH1sArgAl1al1eLeu 159
DB 493 GCTGCGTACAGATGCAACAGGTCACCTTGCGAAAGAAACACATCGCGTATTTCTG 552
QY 160 PheCys1ysG1n1ArgAspLeuLeuProGlnAsnAsnAl1aVal1LeuVal1AlaSerGlyGly 179
DB 553 TTTTGCAGCAGAAAAATTTGCTCTCTCCAGCTTAACGCAATTAATGATGATCTGAGAGT 612
QY 180 Val1AlaSerAsnPheThr11eArgArgAl1eLeuGly11eLeuThraSnAl1aThrGlnCys 199
DB 613 GTTGCAGTAACTTGTATCATCCGAAAGACATTTGAAATTTGCTCAATGCAAGCAGTGC 672

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: November 10, 2005, 16:59:00 ; Search time 886.614 Seconds
(without alignments)
2490.420 Million cell updates/sec

Title: US-10-649-273-2_COPY_148_414
Perfect score: 1385
Sequence: 1 MEAHALTRLRNKNVFPPLV.....DISKEVGASIKVQPKMEI 267

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 9794790 seqs, 413490567 residues

Total number of hits satisfying chosen parameters: 19589580

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-O=/cgsm2_1/USFTO_spool_p/US10649273/runat_02112005_091339_15609/app_query.fasta_1.1429
-DB=Published Applications NA -OPMT=fastcap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bites -START=1 -END=1 -MATRIX=bloum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=plc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10649273 @CGN 1.1 1034 @runat_02112005_091339_15609
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEMEMORY -NEG_SCORES=0 -WAIT_DSPBLOCK=100
-LONGLOG -DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications NA: *
1: /cgsm2_6/ptodata/1/pubpna/US07_PUBCOMB.seq: *
2: /cgsm2_6/ptodata/1/pubpna/PT1_NEW_PUB.seq: *
3: /cgsm2_6/ptodata/1/pubpna/US06_NEW_PUB.seq: *
4: /cgsm2_6/ptodata/1/pubpna/US06_PUBCOMB.seq: *
5: /cgsm2_6/ptodata/1/pubpna/US07_NEW_PUB.seq: *
6: /cgsm2_6/ptodata/1/pubpna/PT01_PUBCOMB.seq: *
7: /cgsm2_6/ptodata/1/pubpna/US08_NEW_PUB.seq: *
8: /cgsm2_6/ptodata/1/pubpna/US08_PUBCOMB.seq: *
9: /cgsm2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq: *
10: /cgsm2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq: *
11: /cgsm2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq: *
12: /cgsm2_6/ptodata/1/pubpna/US09_NEW_PUB.seq: *
13: /cgsm2_6/ptodata/1/pubpna/US09_NEW_PUB.seq: *
14: /cgsm2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq: *
15: /cgsm2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq: *
16: /cgsm2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq: *
17: /cgsm2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq: *
18: /cgsm2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq: *
19: /cgsm2_6/ptodata/1/pubpna/US10G_PUBCOMB.seq: *
20: /cgsm2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq: *
21: /cgsm2_6/ptodata/1/pubpna/US10I_PUBCOMB.seq: *
22: /cgsm2_6/ptodata/1/pubpna/US10J_PUBCOMB.seq: *
23: /cgsm2_6/ptodata/1/pubpna/US10_NEW_PUB.seq: *
24: /cgsm2_6/ptodata/1/pubpna/US10_NEW_PUB.seq: *
25: /cgsm2_6/ptodata/1/pubpna/US11A_PUBCOMB.seq: *
26: /cgsm2_6/ptodata/1/pubpna/US11_NEW_PUB.seq: *
27: /cgsm2_6/ptodata/1/pubpna/US60_NEW_PUB.seq: *
28: /cgsm2_6/ptodata/1/pubpna/US60_PUBCOMB.seq: *

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1385	100.0	1416	18	US-10-120-988-177	Sequence 177, App
2	1385	100.0	1526	15	US-10-067-443-23	Sequence 23, Appl
3	1385	100.0	1526	19	US-10-649-273-23	Sequence 23, Appl
4	1385	100.0	1526	19	US-10-651-722-23	Sequence 23, Appl
5	1385	100.0	2197	15	US-10-067-443-1	Sequence 1, Appl
6	1385	100.0	2197	19	US-10-649-273-1	Sequence 1, Appl
7	1385	100.0	2197	19	US-10-651-722-1	Sequence 1, Appl
8	1385	100.0	2572	22	US-10-480-988-36	Sequence 36, Appl
9	1358.5	98.1	1387	15	US-10-067-443-21	Sequence 21, Appl
10	1358.5	98.1	1387	19	US-10-649-273-21	Sequence 21, Appl
11	1358.5	98.1	1387	19	US-10-651-722-21	Sequence 21, Appl
12	1358.5	98.1	1387	22	US-10-887-553A-1047	Sequence 1047, Ap
13	1348	97.3	1245	15	US-10-012-140-6	Sequence 6, Appl
14	1348	97.3	1245	15	US-10-012-140-4	Sequence 4, Appl
15	1204	86.9	2208	18	US-10-094-749-400	Sequence 400, App
16	1204	86.9	2890	21	US-10-723-860-7447	Sequence 7447, Ap
17	824.5	59.5	14364	15	US-10-067-443-20	Sequence 20, Appl
18	824.5	59.5	14364	19	US-10-649-273-20	Sequence 20, Appl
19	824.5	59.5	14364	19	US-10-651-722-20	Sequence 20, Appl
20	599	43.2	2734	24	US-10-450-763-20426	Sequence 20426, A
21	468	33.8	371	20	US-10-430-201-3118	Sequence 3118, Ap
22	468	33.8	371	20	US-10-430-201-3119	Sequence 3119, Ap
23	422.5	30.5	1917	19	US-10-424-599-66417	Sequence 66417, A
24	415.5	30.0	1601	26	US-11-097-143-34190	Sequence 34190, A
25	415.5	30.0	3656	26	US-11-097-143-34189	Sequence 34189, A
26	359	25.9	1628	20	US-10-437-963-11249	Sequence 11249, A
27	342	24.7	756	15	US-10-081-051-8	Sequence 8, Appl
28	342	24.7	4360	15	US-10-081-051-2	Sequence 2, Appl
29	321.5	23.2	1146	18	US-10-282-122A-14674	Sequence 14674, A
30	314.5	22.7	936	18	US-10-282-122A-8315	Sequence 8315, Ap
31	308.5	22.3	1044	18	US-10-282-122A-26972	Sequence 26972, A
32	308.5	22.3	94750	19	US-10-672-787-38	Sequence 38, Appl
33	304.5	22.0	1032	18	US-10-282-122A-11043	Sequence 11043, A
34	301.5	21.8	1000	20	US-10-343-561-50	Sequence 50, Appl
35	291.5	21.0	1026	9	US-09-815-242-7701	Sequence 7701, Ap
36	291.5	21.0	1026	18	US-10-282-122A-30016	Sequence 30016, A
37	291.5	21.0	1026	24	US-10-958-216-49	Sequence 49, Appl
38	289.5	20.9	1026	24	US-10-958-216-51	Sequence 51, Appl
39	289.5	20.9	1029	9	US-09-815-242-6946	Sequence 6946, Ap
40	289.5	20.9	1029	18	US-10-282-122A-22020	Sequence 22020, A
41	289.5	20.9	1830121	18	US-10-329-670-1	Sequence 1, Appl
42	289.5	20.9	1830121	21	US-10-158-865-1	Sequence 1, Appl
43	289.5	20.9	1830121	24	US-10-961-687-1	Sequence 1, Appl
44	285	20.6	372	21	US-10-357-990-61088	Sequence 61088, A
45	282.5	20.4	9967	24	US-10-795-159-529	Sequence 529, App

ALIGNMENTS

RESULT 1
US-10-120-988-177
; Sequence 177, Application US/10120988
; Publication No. US20030219745A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Goodrich, Ryle
; APPLICANT: Liu, Chenghua
; APPLICANT: Ren, Feiyun
; APPLICANT: Wang, Dunrui
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US20030219745A1 Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 802CON

```

; CURRENT APPLICATION NUMBER: US/10/120,988
; CURRENT FILING DATE: 2002-04-11
; PRIOR APPLICATION NUMBER: 09/774,528
; PRIOR FILING DATE: 2001-01-30
; NUMBER OF SEQ ID NOS: 441
; SOFTWARE: pc FL_genes Version 2.0
; SEQ ID NO 177
; LENGTH: 1416
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (205)..(1305)
US-10-120-988-177

Alignment Scores:
Pred. No.: 1.9e-169      Length: 1416
Score: 1385.00          Matches: 267
Percent Similarity: 100.00%  Conservative: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match: 100.00%      Indels: 0
DB: 18                  Gaps: 0

US-10-649-273-2_COPY_148_414 (1-267) x US-10-120-988-177 (1-1416)
QY      1  MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20
Db      502  ATGAGAGCTCATGACCTTACTATTAGTTGACCAATAAGTGAATTTCTTTTATGTT 561
QY      21  LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
Db      562  CTTTGTGATTTCTGAGAGTCACTGCTGTTGGCATTAGTTCAAGAGATTTCAGATTTTCG 621
QY      41  LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60
Db      622  CTTCTTGAAAGTCTTTGGACATGACACAGGTGACACGCTTGACACAGGTGCAAGGTGCAAGA 681
QY      61  LeuSerLeuIleLysHisProGluCysSerThrMetSerGlyGlyLysAlaIleGluHis 80
Db      682  CTTTCTTTAATAAACAATCCAGAGTCTCCACATGAGTGGGGAAGGCCATGAAACAT 741
QY      81  LeuAlaLysGlnGlyAsnArgPheHisAspIleLysProProLeuHisHisAlaLys 100
Db      742  TTGGCCAAACAGAGAAATAGATTTCATTTTGACATCAAACTCCCTTGATCATGCTAA 801
QY      101  AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120
Db      802  AATTGTGATTTTCTTTTACTGCACTTCAACACGTTACTGATAAATAATATGAAAAAG 861
QY      121  GluLysGluGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaAla 140
Db      862  GAAAAAGAGGAAGGATTTGAGAAAGGGGCAAAATCCGTCTTCAGCAGCAGCAATTCCTG 921
QY      141  ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160
Db      922  ACAGTACAGCACAATGGCATGTATCTTGTGAAAAAGAACACATCGGGCTATTTCTGTT 981
QY      161  CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyVal 180
Db      982  TGTAGCAGAGAGACTTGTATCTCAAAATATATGACATGCTGTGCACTGTGTGTGTGTC 1041
QY      181  AlaSerAsnPheThrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThr 200
Db      1042  GCAAGTAACTTCTATATCCGAGAGCTTGGAAATTTTAAACAAAGCAACAGATGCACT 1101
QY      201  LeuLeuCysProProArgLeuCysThrAspAsnGlyIleMetIleAlaTrpAsnGly 220
Db      1102  TTGTGTGTCTCTCTCCAGACTATGCACTGATTAATGGCATTAATGATTTGCATGGAATGT 1161
QY      221  IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgGlyArg 240
Db      1162  ATTGAAGACTAAGTGTGGCTTGGGCAATTTTACATGACATAGAAAGGATCCGCTATGAA 1221
```

```

QY      241  ProLysCysProLeuGlyValAspIleSerLysGluValGlyGluAlaSerIleLysVal 260
Db      1222  CCAAAATGTCCTCTTGAGTAGATCATATCAAAAGAAAGTTGAGAGAGCTTCCATTAAGA 1281
QY      261  ProGlnLeuLysMetGluIle 267
Db      1282  CCACAATTAAAAATGAGATA 1302

RESULT 2
US-10-067-443-23
; Sequence 23, Application US/10067443
; Publication No. US20030082782A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED I
; TITLE OF INVENTION: SPINAL CORD, MP-1
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 23
; LENGTH: 1526
; TYPE: DNA
; ORGANISM: homo sapiens
US-10-067-443-23

Alignment Scores:
Pred. No.: 2.13e-169      Length: 1526
Score: 1385.00          Matches: 267
Percent Similarity: 100.00%  Conservative: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match: 100.00%      Indels: 0
DB: 15                  Gaps: 0

US-10-649-273-2_COPY_148_414 (1-267) x US-10-067-443-23 (1-1526)
QY      1  MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20
Db      1  ATGAGAGCTCATGACCTTACTATTAGTTGACCAATAAGTGAATTTCTTTTATGTT 60
QY      21  LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
Db      61  CTTTGTGATTTCTGAGAGTCACTGCTGTTGGCATTAGTTCAAGAGATTTCAGATTTTCG 120
QY      41  LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60
Db      121  CTTCTTGAAAGTCTTTGGACATGACACAGGTGACATGCTTACAAAGTGGCAAGAGA 180
QY      61  LeuSerLeuIleLysHisProGluCysSerThrMetSerGlyGlyLysAlaIleGluHis 80
Db      181  CTTTCTTTAATAAACAATCCAGAGTCTCCACATGAGTGGTGGAAAGCCATAGAACAT 240
QY      81  LeuAlaLysGlnGlyAsnArgPheHisAspIleLysProProLeuHisHisAlaLys 100
Db      241  TTGGCCAAACAGGAATATGATTTTCAATTTGACATCAAACTCCCTTGATCATGCTAA 300
QY      101  AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120
Db      301  AATTGTGATTTTCTTTTACTGACTTCAACAGCTTATGATTAATAATATGAAAGAG 360
QY      121  GluLysGluGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaAla 140
Db      361  GAAAAAGAGAGATTTTGAAGAGGGCAAAATCTGTCTTCAAGACAGCATTTGCTGCC 420
QY      141  ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160
Db      421  ACGATACAGACACAATGGCATGTCACTTGTGAAAAAGAACACATCGGGCTATTCGTTT 480
```


US-10-649-273-2_COPY_148_414 (1-267) x US-10-651-722-23 (1-1526)

QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20
DB 1 ATGGAGGCTCATGCACTTACTTATAGTTGACCAATTAAGTAAGATTCTTTTATGTT 60
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
DB 61 CTTTGTATTTCTGAGGCTCATGCTGCTGTTGGCATTAGTTCAAGAGATTTCAGATTTCTG 120
QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60
DB 121 CTTCTTGAAAGTCTTTGGACATGACACAGGTGACATGCTTGACCAAGGTGGCAAGAGA 180
QY 61 LeuSerLeuIleLysHisProGlyCysSerThrMetSerGlyGlyLysAlaIleGlnHis 80
DB 181 CTTTCTTTAATTAACATCCAGAGTCTCCACCATGATGATGGTGGAAAGCCATAGAACAT 240
QY 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 100
DB 241 TTGGCCAAACAGAGAAATAGATTTCATTTTGACATCAAACTCCCTTGATCATGCTAA 300
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120
DB 301 AATTGATATTTTCTTTACTGACCTTCACACCGTTACTGATTAATAATATGAAAAAG 360
QY 121 GluLysGlnGlyGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAla 140
DB 361 GAAAAAGAGAGGATTGAGAGAGGCAAACTGCTTCAGCAGCAGCATTCCTGCC 420
QY 141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160
DB 421 ACAGTACAGCACAACATGCAATGTCATCTTGGAAGAAACACATCGGGCTATTCTGTTT 480
QY 161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyVal 180
DB 481 TGTAAAGCAGAGACTGTGTACTCAAAATATGACATGCTGTGTCATCTGTGTGTGC 540
QY 181 AlaSerAsnPheTyrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThr 200
DB 541 GCAAGTAACTTCTATATCCGAGAGCTCTGGAATTTTAAACAAAGCAACAGTGCCT 600
QY 201 LeuLeuCysProProArgLeuCysThrAspAsnGlyIleMetIleAlaThrPasnGly 220
DB 601 TTGTTGTCTCTCTCCACAGACTATGCACTGATTAATGCAATTATGATGCAATGAT 660
QY 221 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlnGlyIleArgTyrGlu 240
DB 661 ATTGAAGACTACGTCGTGGCTTGGCATTTTACATGACATGAAGGCAATCCGCTATGAA 720
QY 241 ProLysCysProLeuGlyValAspIleSerLysGlnValGlyGluAlaSerIleLysVal 260
DB 721 CCAAAATGCTCTCTTGGAGTAGACATATCAAAAGAAAGTTGGAAGCTTCCATAAAGTA 780
QY 261 ProGlnLeuLysMetGluIle 267
DB 781 CCACATTAATAATGAGATA 801

RESULT 5
US-10-067-443-1
; Sequence 1, Application US/10067443
; Publication No. US20030082782A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10

; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 2197
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (231) .. (1472)
US-10-067-443-1

Alignment Scores:
Pred. No.: 3,67e-169 Length: 2197
Score: 1385.00 Matches: 267
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 15 Gaps: 0

US-10-649-273-2_COPY_148_414 (1-267) x US-10-067-443-1 (1-2197)

QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20
DB 672 ATGGAGGCTCATGCACTTACTTATAGTTGACCAATTAAGTAAGATTCTTTTATGTT 731
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
DB 732 CTTTGTATTTCTGAGGCTCATGCTGCTGTTGGCATTAGTTCAAGAGATTTCAGATTTTCTG 791
QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60
DB 792 CTTCTTGAAAGTCTTTGGACATGACACAGGTGACATCTTGACCAAGGTGGCAAGAGA 851
QY 61 LeuSerLeuIleLysHisProGlyCysSerThrMetSerGlyGlyLysAlaIleGlnHis 80
DB 852 CTTTCTTTAATAAACATCCAGAGTCTCCACCATGATGATGGAGGCAATAGAACAT 911
QY 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 100
DB 912 TTGGCCAAACAGAGAAATAGATTTCATTTTGACATCAAACTCCCTTGATCATGCTAA 971
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120
DB 972 AATTGATATTTTCTTTACTGCACTTCAACAGTACATGATTAATAATGAAAAAG 1031
QY 121 GluLysGlnGlyGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAla 140
DB 1032 GAAAAAGAGAGGTAATTGAGAGGGCAAACTCTGCTTCAGCAGCAGCATTCGCTGCC 1091
QY 141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160
DB 1092 ACAGTACAGCACAACATGCAATGTCATCTTGGAAGAAACACATGGGCTATTCTGTTT 1151
QY 161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyVal 180
DB 1152 TGTAAAGCAGAGACTGTGTACTCAAAATATGCACTGCTGTCATCTGTGTGTGC 1211
QY 181 AlaSerAsnPheTyrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThr 200
DB 1212 GCAAGTAACTTCTATATCCGAGAGCTCTGGAATTTTAAACAAACCAACACAGTGCCT 1271
QY 201 LeuLeuCysProProArgLeuCysThrAspAsnGlyIleMetIleAlaThrPasnGly 220
DB 1272 TTGTTGTCTCTCTCCACAGACTATGCACTGATTAATGCAATTATGATGCAATGAT 1331
QY 221 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlnGlyIleArgTyrGlu 240
DB 1332 ATTGAAGACTACGTCGTGGCTTGGCATTTTACATGACATGAAGGCAATCCGCTATGAA 1391
QY 241 ProLysCysProLeuGlyValAspIleSerLysGlnValGlyGluAlaSerIleLysVal 260
DB 1392 CCAAAATGCTCTCTTGGAGTAGACATATCAAAAGAAAGTTGGAAGGCTTCCATAAAGTA 1451

QY 261 ProGlnLeuLysMetGluile 267
DB 1452 CCACAATTAAAAATGGAGATA 1472

RESULT 6
US-10-649-273-1
; Sequence 1, Application US/10649273
; Publication No. US20040043407A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1
; FILE REFERENCE: D0073 CMT
; CURRENT APPLICATION NUMBER: US/10/649,273
; CURRENT FILING DATE: 2003-08-27
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 2197
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (231)..(1472)
US-10-649-273-1

Alignment Scores:
Pred. No.: 3,67e-169 Length: 2197
Score: 1385.00 Matches: 267
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 19 Gaps: 0

US-10-649-273-2_COPY_148_414 (1-267) x US-10-649-273-1 (1-2197)

QY 1 MetGlnAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20
DB 672 ATGGAGGCTCATGCACTTACTATTAGGTTGACCAATTAAGTAAGATTCTTTTAACTT 731
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
DB 732 CTTTGAATTTCTGGAGGTCACTGCTGTTGGCAATTGTTCAAGAGTTTCAGATTTTCTG 791
QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60
DB 792 CTTCTTGGAAAGCTTTTGGACATGACACAGGTGACATGCTTGCACAGGTGGCAAGAGA 851
QY 61 LeuSerLeuIleLysHisProGluCysSerThrMetSerGlyGlyLysAlaIleGluHis 80
DB 852 CTTTCTTTAATAAATCATCCAGAGTGTCCACCACTGATGTTGGTGGAAAGCCATGAACAT 911
QY 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProPheLeuHisAlaLys 100
DB 912 TTGGCCAAACAAAGCAATATGATTCATTGACATCAACCTCCCTTGCATCAGCTTAA 971
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120
DB 972 AATGTGATTTTCTTTACTGACCTTCAACAGTTACTGATTAATAATATATAAAG 1031
QY 121 GluLysGluGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaAla 140
DB 1032 GAAAAAGAGAGAGATTTAGAGAGGGGCAATCTGCTTTCAGACAGACATTTCTGCC 1091
QY 141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160
DB 1092 ACAGTACAGACACATATGCAATCTTGTGTAAAGAACATCGGGCATATCTGTTT 1151

QY 161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyVal 180
DB 1152 TGTAAAGCAGAGAGACTTGTACTTCAATAATATGACGTACTGTTCCATCTGCTGTGTC 1211
QY 161 AlaSerAspPheThrIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThr 200
DB 1212 GCAAGTAACTTTATATATCCGACAGCTCTGAAATTTTAAACAAACCAACACAGTCACT 1271
QY 201 LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaTrpAsnGly 220
DB 1272 TTGTTGTGTCCTCCCTCCCAAGACTGATGCACTGATTAATGCGATTATGATGATGATG 1331
QY 221 IleGluValGluArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgGlyVal 240
DB 1332 ATTTGAAGACTACGTGCTGGCTTGGGCAATTTTACATGACATGAAAGCATCCGCTATGAA 1391
QY 241 ProLysCysProLeuGlyValAlaAspIleSerLysGluValGlyGluAlaSerIleLysVal 260
DB 1392 CCAAAATGCTCTTGTGAGTAGACATATCAAAAGAAAGTTGAGAGACCTTCATAAAGTA 1451
QY 261 ProGlnLeuLysMetGluile 267
DB 1452 CCACAATTAAAAATGGAGATA 1472

RESULT 7
US-10-651-722-1
; Sequence 1, Application US/10651722
; Publication No. US20040048302A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1
; FILE REFERENCE: D0073 DIV
; CURRENT APPLICATION NUMBER: US/10/651,722
; CURRENT FILING DATE: 2003-08-29
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 2197
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (231)..(1472)
US-10-651-722-1

Alignment Scores:
Pred. No.: 3,67e-169 Length: 2197
Score: 1385.00 Matches: 267
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 19 Gaps: 0

US-10-649-273-2_COPY_148_414 (1-267) x US-10-651-722-1 (1-2197)

QY 1 MetGlnAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20
DB 672 ATGGAGGCTCATGCACTTACTATTAGGTTGACCAATTAAGTAAGATTCTTTTAACTT 731
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
DB 732 CTTTGAATTTCTGGAGGTCACTGCTGTTGGCAATTGTTCAAGAGTTTCAGATTTTCTG 791
QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60
DB 792 CTTCTTGGAAAGCTTTTGGACATGACACAGGTGACATGCTTGCACAGGTGGCAAGAGA 851

```
QY 61 LeuSerLeuIleYshISProGluCySerThrMetSerGlyGlyValaIleGluHis 80
Db 852 CTTTCTTTAATAAATCAATCCAGAGGTCCACCATGATGGTGGGAAAGCCATAGAAAT 911
QY 81 LeuAlaIleGlnGlyYAsnArgPheHisPheAspIleYleYProProLeuHisHisAlaIys 100
Db 912 TTGGCCAAACAGAGAAATGATTCATTTTTCATGCAAACTCCCTTCATCATCTAA 971
QY 101 AsnCyAspPheSerPheThrGlyLeuGlnHisValThraAspYleIleMetYshIys 120
Db 972 AATTGTGATTTTCTTTTACTGAGACTTCAACACGTTACTGATTAATTAATGAAAAAG 1031
QY 121 GluIleGlnGluGlyIleGluIleGlyGlnIleLeuSerSerAlaAlaAspIleAla 140
Db 1032 GAAAAAGAGGAGGATGAGAAAGGGCAAAATCCTGCTTCAGCAGCAGACATTCCTGCC 1091
QY 141 ThrValGlnHisThrMetAlaCyHisIleuValIysaArgThrHisArgAlaIleLeuPhe 160
Db 1092 ACAGTACAGCACACAAATGGCATGTCATCTTGTGAAAAACACATCGGGCTAATCTGTTT 1151
QY 161 CysIlySGlnArgAspLeuLeuProGlnAsnAsnAlaValIleuValAlaSerGlyVal 180
Db 1152 TGTAAAGCAGAGAGACTTGTACCTCCAAATATATGACAGTACTGTCATCTGGTGTC 1211
QY 181 AlaSerAsnPhetyrIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThr 200
Db 1212 GCAAGTAACTTCTATATCGCAGAGCTCTGGAAATTTTAAACAAACGCAACACAGTCACT 1271
QY 201 LeuLeuCySPProProProArgLeuCySPThrAspAsnGlyIleMetIleAlaTrpAsnGly 220
Db 1272 TTGTGTGTCTCTCCCTCCAGCATATGACATGCTATATATGATGATGATGATGATGAT 1331
QY 221 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTrpGlu 240
Db 1332 ATTAAGAAACACGAGCTGGCTGGCAATTTTACATGACATGAGAGGCTCCGCTATGAA 1391
QY 241 ProIysCySPProLeuGlyValAspIleSerIysGluValGlyGluAlaSerIleYshVal 260
Db 1392 CCAAAATCTCTCTTGGAGTACATATCCAAAGAGTGGAGAACTTCCATTAAGA 1451
QY 261 ProGlnLeuIysMetGluIle 267
Db 1452 CCACAAATTAATAATGAGATA 1472

RESULT 8
US-10-480-988-36
; Sequence 36, Application US/10480988
; Publication No. US20050069877A1
; GENERAL INFORMATION:
; APPLICANT: GANDHI, Ameena R.; KABLE, Amy E.;
; APPLICANT: SARANAKAR, Anita; HARALIA, April J.A.;
; APPLICANT: TRAN, Bao; DUGGAN, Brendan M.;
; APPLICANT: WARREN, Bridget A.; ISON, Craig H.;
; APPLICANT: HONCHELL, Cynthia D.; NGUYEN, Daniel B.;
; APPLICANT: LU, Dzung Aina M.; LEE, Ernestine A.;
; APPLICANT: YUE, Henry; FORSYTHE, Ian J.;
; APPLICANT: BARROSO, Ines; RAMKUMAR, Jayalaxmi;
; APPLICANT: GRIFFIN, Jennifer A.; LI, Joana X.;
; APPLICANT: YANG, Junning; THANGAVELU, Kavitha;
; APPLICANT: GIETZEN, Kimberly J.; DING, Li;
; APPLICANT: BAUGHN, Mariah R.; BOROMSKY, Mark L.;
; APPLICANT: YAO, Monique G.; CHAMLA, Narinder K.;
; APPLICANT: MASON, Patricia M.; GURURAJAN, Rajagopal;
; APPLICANT: LEE, Sally; BECHA, Shanya D.;
; APPLICANT: ELIOTT, Vicki S.; LUO, Wen;
; APPLICANT: SPRAGUE, William W.; TANG, Y. Tom;
; APPLICANT: LU, Yan; ZEBARADIAN, Yeganeh
; TITLE OF INVENTION: PROTEIN MODIFICATION AND MAINTENANCE MOLECULES
; FILE REFERENCE: PF-1040 USN
; CURRENT APPLICATION NUMBER: US/10/480,988
; CURRENT FILING DATE: 2003-12-16
; PRIOR APPLICATION NUMBER: PCT/US02/19360
```

```
; PRIOR FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: US 60/300,508
; PRIOR FILING DATE: 2001-06-22
; PRIOR APPLICATION NUMBER: US 60/303,445
; PRIOR FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US 60/305,405
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/311,442
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: US 60/314,821
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/315,992
; PRIOR FILING DATE: 2001-08-29
; PRIOR APPLICATION NUMBER: US 60/378,205
; PRIOR FILING DATE: 2002-05-03
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: PERL Program
; SEQ ID NO 36
; LENGTH: 2572
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No: 7632424CB1
; US-10-480-988-36

Alignment Scores:
Pred. No.: 4 65e-169 Length: 2572
Score: 1385.00 Matches: 267
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 22 Gaps: 0

US-10-649-273-2_COPY_148_414 (1-267) x US-10-480-988-36 (1-2572)
QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnIleValGluPheProPheLeuVal 20
Db 585 ATGGAGGCTCAGACACTTATCTATTTAGCTTGACCAATTAAGTGAATTTCTTTTATGTT 644
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
Db 645 CTTTGTGATTTCTGGAGGTCACTGCTGTGGCATTTGTTCAAGAGATTCAGATTTTCTG 704
QY 41 LeuLeuGlyIleSerLeuAspIleAlaProGlyAspMetLeuAspIysValAlaArgArg 60
Db 705 CTTTGTGAAAGCTTTTGGACATAGCACCAAGGTGACATCTTGACAAAGTGGCAGAGAGA 764
QY 61 LeuSerLeuIleYshISProGluCySerThrMetSerGlyGlyYshAlaIleGluHis 80
Db 765 CTTTCTTTAATAAATCAATCCAGAGGTCCACCATGATGGTGGGAAAGCCATAGAAAT 824
QY 81 LeuAlaIleGlnGlyYAsnArgPheHisPheAspIleYleYProProLeuHisHisAlaIys 100
Db 825 TTGGCCAAACAGAGAAATGATTCATTTTTCATGCAAACTCCCTTCATCATCTAA 884
QY 101 AsnCyAspPheSerPheThrGlyLeuGlnHisValThraAspYleIleMetYshIys 120
Db 885 AATTGTGATTTTCTTTTACTGAGACTTCAACACGTTACTGATTAATTAATGAAAAAG 944
QY 121 GluIleGlnGluGlyIleGluIleGlyGlnIleLeuSerSerAlaAlaAspIleAla 140
Db 945 GAAAAAGAGGAGGATGAGAAAGGGCAAAATCCTGCTTCAGCAGCAGACATTCCTGCC 1004
QY 141 ThrValGlnHisThrMetAlaCyHisIleuValIysaArgThrHisArgAlaIleLeuPhe 160
Db 1005 ACAGTACAGCACACAAATGGCATGTCATCTTGTGAAAAACACATCGGGCTAATCTGTTT 1064
QY 161 CysIlySGlnArgAspLeuLeuProGlnAsnAsnAlaValIleuValAlaSerGlyVal 180
Db 1065 TGTAAAGCAGAGAGACTTGTATCTCCAAATATATGACAGTACTGTCATCTGGTGTC 1124
QY 181 AlaSerAsnPhetyrIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThr 200
```

Db 1125 GCAAGTAATTCTTATATCCGACAGCTCTGGAAATTTTAAACAAACCAACACAGTCCACT 1184
Qy 201 LeuLeuCyProProProArgLeuCyThrAspAsnGlyIleMetIleAlaTrpAsnGly 220
Db 1185 TTGTTGTGTCCCTCCGACAGTATGCACTGATTAATGGCATTATGATTCATGGAATGCT 1244
Qy 221 TleGluAlaGluAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyGlu 240
Db 1245 ATTGAAGACTAGCTGCTGGCTTGGCATTTTACATGACATGAAGGCAATCCGCTATGAA 1304
Qy 241 ProLysCysProLeuGlyValAlaSerLysGluValGlyGluAlaSerIleLysVal 260
Db 1305 CCAAAATGTCCTCTTGAGTAGACATATCAAAAGAGTTGAGAAAGCTTCCATAAAGTA 1364
Qy 261 ProGlnLeuLysMetGluIle 267
Db 1365 CCACATTTAAATGAGATA 1385

RESULT 9
US-10-667-443-21
; Sequence 21, Application US/10067443
; Publication No. US20030082782A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 21
; LENGTH: 1387
; TYPE: DNA
; ORGANISM: homo sapiens
US-10-667-443-21

Alignment Scores:
Pred. No.: 5,276-166 Length: 1387
Score: 1358.50 Matches: 266
Percent Similarity: 91.10% Conservative: 0
Best Local Similarity: 91.10% Mismatches: 1
Query Match: 98.09% Indels: 25
Gaps: 1
US-10-649-273-2_COPY_148_414 (1-267) x US-10-667-443-21 (1-1387)
Qy 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20
Db 465 ATGAGGCTCATGACCTTACTATTAGTTGACCAATTAAGTGAATTCCTTTTATAGTT 524
Qy 21 LeuLeuLysSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
Db 525 CTTTGATTCTCGAGGTCACGCTGCTGTTGGCATTAGTTCAAGAGGTTTCAATTTTCTG 584
Qy 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60
Db 585 CTTCTTGAAAGTCTTTGACATGACCAAGGTCAGCTTGACCAAGGTCGCAAGAA 644
Qy 61 LeuSerLeuIleLysHisAspGlnCysSerThrMetSerGlyGlyLysAlaIleGluHis 80
Db 645 CTTTCTTAAATAAACATCCAGAGTCTCCACCATGAGTGGGAAAGCCATAGAGCAT 704
Qy 81 LeuAlaLysGlnGlyAsnArgPheHisAspAspIleLysProProLeuHisHisAlaLys 100
Db 705 TTGGCCAAACAAAGAAATGATTTTCATTTTGCATCAAAACCTCCCTGCATCATCTAAA 764
Qy 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleLeuLysLys 120

Db 765 AATTGGAATTTTCTTTTACTGACCTTCAACAGTTATCGATTAATAATTAATGAAGAAAG 824
Qy 121 GluLysGluGluGlyIle----- 126
Db 825 GAAAAAGAGAGAGTATATTTCTAATTAGTAAGTTGAACAGATTAATATCTCGATTG 884
Qy 127 -----GluLysGlyGlnIleLeuSerSerAla 135
Db 885 TGCCATAAATAAGCTGCTCATTTCTGACAGTATAGAGAGGGCAATCTGCTTCAGCA 944
Qy 136 AlaAspIleAlaAlaThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHis 155
Db 945 GCGAGCATTTGCTGCCACAGTACGACACAAATGGCATGTTCATCTTGAAAGAACACAT 1004
Qy 156 ArgAlaIleLeuPheCysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuVal 175
Db 1005 CGGGCTATTCTGTTTGTAAAGCAGAGACTGTGTACCTCAAAATTAATGACGACTGTT 1064
Qy 176 AlaSerGlyGlyValAlaSerAsnPheTyIleArgArgAlaLeuGluIleLeuThrAsn 195
Db 1065 GCATCTGAGTGCTGTCAGAGTAACTTATATCCGACAGCTCTGGAATTTTAAACAAAC 1124
Qy 196 AlaThrGlnCysThrLeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMet 215
Db 1125 GCAACACAGTGCACCTTGTGTCTCTCCACAGACTAGCACTGATATATGCAATTATG 1184
Qy 216 AlaThrAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlu 235
Db 1185 ATTCATGTAAGTAATGTAATTAAGACTACGTGCTGGCATTTTACATGACATAGAA 1244
Qy 236 GlyIleArgTyGlnProLysCysProLeuGlyValAspIleSerLysGluValGlyGlu 255
Db 1245 GGCATCCGCTATGAACCAAAATGTCTCTTGAGTAGACATATCAAAAGAGTTGAGAA 1304
Qy 256 AlaSerIleLysValProGlnLeuLysMetGluIle 267
Db 1305 GCTTCATTAAGTACCAATTAATAATGAGATA 1340

RESULT 10
US-10-649-273-21
; Sequence 21, Application US/10649273
; Publication No. US20040043407A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1
; FILE REFERENCE: D0073 CNT
; CURRENT APPLICATION NUMBER: US/10/649,273
; PRIOR FILING DATE: 2003-08-27
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 21
; LENGTH: 1387
; TYPE: DNA
; ORGANISM: homo sapiens
US-10-649-273-21

Alignment Scores:
Pred. No.: 5,276-166 Length: 1387
Score: 1358.50 Matches: 266
Percent Similarity: 91.10% Conservative: 0
Best Local Similarity: 91.10% Mismatches: 1
Query Match: 98.09% Indels: 25
Gaps: 1
US-10-649-273-2_COPY_148_414 (1-267) x US-10-649-273-21 (1-1387)

QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20
DB 465 ATGAGGCTCATGCACTTACTATTAGTGTGACCAATTAAGTGAATTCCTTTTAGTT 524
QY 21 LeuLeuIleSerGlyValHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
DB 525 CTTTGTATTTCTGGAGGTCACTGTCTGTGGCATTAGTTCAGAGGATTCAGATTTTCTG 584
QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60
DB 585 CTCTTGAAAGTCTTTTGACATGACACCAAGGTGACATCTTGACAAAGGTGGCAAGAGA 644
QY 61 LeuSerLeuIleLysHisProGluCysSerThrMetSerGlyLysAlaIleGluHis 80
DB 645 CTTTCTTTAATAAATCAATCCAGAGTCTCCACATGAGTGTGGGAAAGCCATAGAGCAT 704
QY 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 100
DB 705 TTGGCCAAACAAAGAAATAGATTTCATTTTGACATCAAACTCCCTTGATCATGCTAA 764
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120
DB 765 AATTGTGATTTTCTTTTACTGAGCTTCAACAGTTACTGATTAATAATTAATGAAGAAAG 824
QY 121 GluLysGlnGluGlyIle-----GluLysGlnGlnIleLeuSerSerAla 126
DB 825 GAAAGAGGAAAGATATTTCTATTAAGTTGAACAGATAAATATTCCTGATTTG 884
QY 127 -----GluLysGlnGlnIleLeuSerSerAla 135
DB 885 TGCCCTAAATATAGCTCATTTCTGACAGGTATGAGAGGAGGCAAAATCCTGCTTCAGCA 944
QY 136 AlaAspIleAlaIleAlaThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHis 155
DB 945 GCAGCATTTGCTGCACAGTACAGACACAAATGGCATGTCATCTGTGAAAGAAAGACAT 1004
QY 156 ArgAlaIleLeuPheCysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuVal 175
DB 1005 CGGGCTATTTCTTTTGTGAAGAGAGACCTTGTACCTCAAAATTAATGACATCTGTT 1064
QY 176 AlaSerGlyGlyValAlaSerAsnPheTyrlleArgArgAlaLeuGlnIleLeuThrAsn 195
DB 1065 GCATCTGGAGGTGTCGCAAGTAACTTCTATATCCGAGAGCTCTGGAATTTTAAACAAC 1124
QY 196 AlaThrGlnCysThrLeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMet 215
DB 1125 GCAACACAGTGCATTTGTGTGTCTCTCCAGACTATGCACTGATATGCGCATTTATG 1184
QY 216 IleAlaIlePheAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlu 235
DB 1185 ATTGATGAGAAATGATTAAGAAAGACTACGTGTGCTTGCGCATTTTACATGACATAGAA 1244
QY 236 GlyIleAlaGlyArgLysProLysCysProLeuGlnGlyValAspIleSerLysGlnValGlyGlu 255
DB 1245 GGCATCCCGCTATGAAACCAAAATGCTCTCTTGAGATGACATATCAAAAGAAAGTTGAGAA 1304
QY 256 AlaSerIleLysValProGlnLeuLysMetGluIle 267
DB 1305 GCTTCATTAATAAGTACCAATTAATAATGAGAGATA 1340

RESULT 11
US-10-651-722-21
; Sequence 21, Application US/10651722
; Publication No. US20040048302A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1
; FILE REFERENCE: D0073 DIV
; CURRENT APPLICATION NUMBER: US/10/651,722
; CURRENT FILING DATE: 2003-08-29
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443

QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20
DB 465 ATGAGGCTCATGCACTTACTATTAGTGTGACCAATTAAGTGAATTCCTTTTAGTT 524
QY 21 LeuLeuIleSerGlyValHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
DB 525 CTTTGTATTTCTGGAGGTCACTGTCTGTGGCATTAGTTCAGAGGATTCAGATTTTCTG 584
QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60
DB 585 CTCTTGAAAGTCTTTTGACATGACACCAAGGTGACATCTTGACAAAGGTGGCAAGAGA 644
QY 61 LeuSerLeuIleLysHisProGluCysSerThrMetSerGlyLysAlaIleGluHis 80
DB 645 CTTTCTTTAATAAATCAATCCAGAGTCTCCACATGAGTGTGGGAAAGCCATAGAGCAT 704
QY 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 100
DB 705 TTGGCCAAACAAAGAAATAGATTTCATTTTGACATCAAACTCCCTTGATCATGCTAA 764
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120
DB 765 AATTGTGATTTTCTTTTACTGAGCTTCAACAGTTACTGATTAATAATTAATGAAGAAAG 824
QY 121 GluLysGlnGluGlyIle-----GluLysGlnGlnIleLeuSerSerAla 126
DB 825 GAAAGAGGAAAGTATATTTCTAATTAGTAAAGTTGAACAGATAAATATTCCTGATTTG 884
QY 127 -----GluLysGlnGlnIleLeuSerSerAla 135
DB 885 TGCCCTAAATATAGCTCATTTCTGACAGGTATGAGAGGAGGCAAAATCCTGCTTCAGCA 944
QY 136 AlaAspIleAlaIleAlaThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHis 155
DB 945 GCAGCATTTGCTGCACAGTACAGACACAAATGGCATGTCATCTGTGAAAGAAAGACAT 1004
QY 196 AlaThrGlnCysThrLeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMet 215
DB 1125 GCAACACAGTGCATTTGTGTGTCTCTCCAGACTATGCACTGATATGCGCATTTATG 1184
QY 216 IleAlaIlePheAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlu 235
DB 1185 ATTGATGAGAAATGATTAAGAAAGACTACGTGTGCTTGCGCATTTTACATGACATAGAA 1244

PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 21
; LENGTH: 1387
; TYPE: DNA
; ORGANISM: homo sapiens
US-10-651-722-21

Alignment Scores:
Pred. No.: 5,27e-166 Length: 1387
Score: 1358.50 Matches: 266
Percent Similarity: 91.10% Conservative: 0
Best Local Similarity: 91.10% Mismatches: 1
Query Match: 98.09% Indels: 25
DB: 19 Gaps: 1

US-10-649-273-2_COPY_148_414 (1-267) x US-10-651-722-21 (1-1387)

```
QY 236 GYIIEARGYRGLUPROLYSCYSPROLEUGLYVALASPILESERLYSGIUVAlGlygu 255
DB 1245 GGCATCCGCTATGAAACCAAAATGCTCTTGAGTAGACATATCAAAAAGTTGGAGAA 1304
QY 256 ALaSerIleLySValProGlnLeuLySmetGluile 267
DB 1305 GCTTCCATTAAGTACCAATTAATAAATGGAGATA 1340
RESULT 12
US-10-887-553A-1047
; Sequence 1047, Application US/10887553A
; Publication No. US20050085436A1
; GENERAL INFORMATION:
; APPLICANT: Garza, Dan
; APPLICANT: Li, Hao
; TITLE OF INVENTION: Method to treat conditions associated
; TITLE OF INVENTION: with insulin signalling dysregulation
; FILE REFERENCE: 4-33262
; CURRENT APPLICATION NUMBER: US/10/887, 553A
; PRIOR FILING DATE: 2004-07-08
; PRIOR APPLICATION NUMBER: 60/485,883
; PRIOR FILING DATE: 2003-08-07
; NUMBER OF SEQ ID NOS: 1208
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1047
; LENGTH: 1387
; TYPE: DNA
; ORGANISM: human
US-10-887-553A-1047
Alignment Scores:
Pred. No.: 5,27e-166 Length: 1387
Score: 1358.50 Matches: 266
Percent Similarity: 91.10% Conservative: 0
Best Local Similarity: 91.10% Mismatches: 1
Query Match: 98.09% Indels: 25
DB: Gaps: 1
US-10-649-273-2_COPY_148_414 (1-267) x US-10-887-553A-1047 (1-1387)
QY 1 MetGlnAlaHisAlaLeuThrIleArgLeuThrAsnLySValGluPheProPheLeuVal 20
DB 465 ATGAGAGCTCATGACCTACTATTAGGTTGACCAATTAAGTAGAATTTCTTTTATGTT 524
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
DB 525 CTTTGTATTTCTGGAGGTCACGTCTGTGTGTCATTAGTTCAGAGAGTTTCAGATTTCGTG 584
QY 41 LeuLeuGlyLySserLeuAspIleAlaProGlyAspMetLeuAspLySValAlaArgArg 60
DB 585 CTTCTTGGAAGGCTTTTGACATAGACACAGGTGACATGCTTGACAGGTGGCAAGAGA 644
QY 61 LeuSerLeuIleLySHisProGlyCysSerThrMetSerGlyGlyValAlaIleGluHis 80
DB 645 CTTTCTTTAATAAACAATCCAGAGTGTCCACACATGAGTGGGAAGGCATAGAGCAT 704
QY 81 LeuAlaLySglnGlyAsnArgPheHisPheAspIleLySProProLeuHisHisAlaLyS 100
DB 705 TTGGCCCAACAAAGAAATAGATTTCATTTCATCAATCAAACTCCCTTCATCATGCTTAA 764
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLySleileMetLySValS 120
DB 765 AATTGTGATTTTCTTTTACTGACCTTCACACGTTACTGATTAATAATATGAAAAAG 824
QY 121 GlnLySglnGlyGlyLeu----- 126
DB 825 GAAAAAAGAGAGCTATATTCTAATTAGTAAGTGAACAGATAAATATTCCTGAGATTG 884
QY 127 -----GlnLySglnGlnIleLeuSerSerAla 135
DB 885 TGCCATAAAATAGCTGCTCATTTCTGACAGTATGAGAGGGAATATCTCTGCTTCAACA 944
QY 136 AlaAspIleAlaIleThrValGlnHisThrMetAlaCysHisIleLeuValLySArgThrHis 155
```

```
DB 945 GCAGACATTGCTGCCACAGTACAGACACAATGGCATGTCACTTGTGAAAAACAACAT 1004
QY 156 ArgAlaIleLeuPheCysLySglnArgAspLeuLeuProGlnAsnAsnAlaValLeuVal 175
DB 1005 CCGGCTATTTCTGTTTGTGTAAGAGAGAGACTGTTACCTCAAAATATATGACATGCTGTT 1064
QY 176 ALaSerGlyGlyValAlaSerAsnPheThrIleArgArgAlaLeuGlnIleLeuThrAsn 195
DB 1065 GCATCTGGTGTGTGCGAAAGTAACTTCTATATCCGACAGAGCTCTGGAAATTTTAAACAAC 1124
QY 196 AlaThrGlnCysThrLeuLeuCysProProProAlaGlyLeuCysThrAspAsnGlyIleMet 215
DB 1125 GCACACAGTGCACCTTGTGTGTCCTCTCCCAACTATGCACTGATATATGCAATTATG 1184
QY 216 IleAlaTPAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlu 235
DB 1185 ATTGATGATGATGATGATTAAGAACTACGTGCTGGCTGGCAATTTTACATATACATAGAA 1244
QY 236 GYIIEARGYRGLUPROLYSCYSPROLEUGLYVALASPILESERLYSGIUVAlGlygu 255
DB 1245 GGCATCCGCTATGAAACCAAAATGCTCTTGAGTAGACATATCAAAAAGTTGGAGAA 1304
QY 256 ALaSerIleLySValProGlnLeuLySmetGluile 267
DB 1305 GCTTCCATTAAGTACCAATTAATAAATGGAGATA 1340
```

```
RESULT 13
US-10-012-140-6
; Sequence 6, Application US/10012140
; Publication No. US20030009017A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Leiby, Kevin R.
; APPLICANT: Kapsler-Libermann, Rosana
; APPLICANT: Glucksmann, Maria A.
; TITLE OF INVENTION: 38650, 28472, 5495, 65507, 81588, AND
; TITLE OF INVENTION: 14354 METHODS AND COMPOSITIONS OF HUMAN PROTEINS AND USBS
; FILE REFERENCE: 381552004900
; CURRENT APPLICATION NUMBER: US/10/012,140
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: 60/246,768
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/246,772
; PRIOR FILING DATE: 2000-11-08
; PRIOR FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 1245
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-012-140-6
```

Alignment Scores:

```
Pred. No.: 1,05e-164 Length: 1245
Score: 1348.00 Matches: 260
Percent Similarity: 98.50% Conservative: 3
Best Local Similarity: 97.38% Mismatches: 4
Query Match: 97.33% Indels: 0
DB: Gaps: 0
```

```
US-10-649-273-2_COPY_148_414 (1-267) x US-10-012-140-6 (1-1245)
```

```
QY 1 MetGlnAlaHisAlaLeuThrIleArgLeuThrAsnLySValGluPheProPheLeuVal 20
DB 442 ATGAGAGCTCATGACCTACTATTAGGTTGACCAATTAAGTGAATTTCTTTTATGTT 501
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
DB 502 CTTTGTATTTCTGGAGGTCACGTCTGTGTGCAATTAGTTCAGAGAGTTTCATTTTCTG 561
```

```
Qy 41 LeuLeuGlyysSerLeuAsp11eAlaProGlyAspMetLeuAspLysVal1AaArg 60
Db 562 CTTCTTGAAAGTCTTTGACATAGCACAGGTGACATGCTTGACAAAGTGGCAAGAAGA 621
Qy 61 LeuSerLeu1eLysHisProGlyuCySerThrMetSerGlyLysVal1AaLeuHis 80
Db 622 CTTCTTTAAATAAACATCCAGAGGCTCCACCAAGAGGGGAGAAAGCCCTAGAACAT 681
Qy 81 LeuAlaLysGlnGlyAsnArgPheHisPheAsp11eLysProProLeuHisHisAlaLys 100
Db 682 TTGGCCAAACAGGAATAGATTTCATTTTGCATCAAACTCCCTGCATCATGCTTAA 741
Qy 101 AsnCyAspPheSerPheThrGlyLeuGlnHisVal1ThrAspLys1eLleMetLysLys 120
Db 742 AATTGTGATTTTCTTTACTGAGACTTCAACGCTTACGTAATAAATAATAGAAAACAG 801
Qy 121 GluLysGlnGlyLys1eLysGlyLysGln1eLeuSerSerAlaAlaAsp11eAla 140
Db 802 AAACAGAGGAAGGATAGAGAGGGGCAAACTGCTTTCAGCAGCAGACATTTGCTGCC 861
Qy 141 ThrValGlnHisThrMetAlaCyHisLeuValLysArgThrHisArgAla1eLeuPhe 160
Db 862 ACAGTACAGCACACATGCGCATGTCATTTGAAAAGAACACATCGGGCTATTCGT 921
Qy 161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaVal1LeuVal1AaSerGlyLysVal 180
Db 922 TGTAAAGCAGAGAGACTTGTACTTCAAAATATGACATGCTGCTGCTGCTGCTGCT 981
Qy 181 AlaSerAsnPheTyrlleArgArgAlaLeuGln1eLeuThrAsnAlaThrGlnCysThr 200
Db 982 GCAAGTAACTTCTATATCCGAGAGCTGGAATTTTAAACAAGCAACACAGTGCAC 1041
Qy 201 LeuLeuCyProProProArgLeuCySerThrAspAsnGlyLysMetLleAla1eLeuGly 220
Db 1042 TTGTGTGTCTCTCCCTCCAGACTATGCACTGATTAATGAGATTAATGATGAGATGCT 1101
Qy 221 IleGluArgLeuArgAlaGlyLeuGlyLys1eLeuHisAsp11eLysGlyLys1eLeuArg 240
Db 1102 ATTGAAGACTACGCTGCTGGCTTGGCATTTTACATGACATGAAAGGATCCGCTATGAA 1161
Qy 241 ProLysCyProLeuGlyValAsp11eSerLysGlnVal1GlyLys1eLeuSerLysVal 260
Db 1162 CCAAAATGCTCTCTTGAGATGACATATCAAAAGAAAGTTGAGAAAGCTTCCATTAAGTA 1221
Qy 261 ProGlnLeuLysMetGlu1e 267
Db 1222 CCACATTAATAAATGAGATA 1242

RESULT 14
US-10-012-140-4
; Sequence 4, Application US/10012140
; Publication No. US20030009017A1
; GENERAL INFORMATION:
; APPLICANT: Leiby, Kevin R.
; APPLICANT: Kapeller-Libermann, Rosana
; TITLE OF INVENTION: 38650, 28472, 5495, 65507, 81588, AND
; TITLE OF INVENTION: 14354 METHODS AND COMPOSITIONS OF HUMAN PROTEINS AND USES
; FILE REFERENCE: 381552004900
; CURRENT APPLICATION NUMBER: US/10/012,140
; CURRENT FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: 60/246,768
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/246,772
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/249,185
; PRIOR FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1820
; TYPE: DNA
```

```
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (146) ... (1390)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1) ... (1820)
; OTHER INFORMATION: n = A,T,C or G
US-10-012-140-4

Alignment Scores:
Score: 1,85e-164 Length: 1820
Percent Similarity: 1348.00 Matches: 260
Best Local Similarity: 98.508 Conservative: 3
Query Match: 97.388 Mismatches: 4
DB: 97.338 Indels: 0
Gaps: 0

US-10-649-273-2_COPY_148_414 (1-267) x US-10-012-140-4 (1-1820)

Qy 1 MetGluAlaHisAlaLeuThr11eArgLeuThrAspLysVal1GluPheProPheLeuVal 20
Db 587 ATGAGGCTCATAGCACTTACTTATTAAGTTGACCAATTAAGTAATTTCTTTTATGTT 646
Qy 21 LeuLeu1eSerGlyLysHisCysLeuLeuAlaVal1GlnGlyVal1SerAspPheLeu 40
Db 647 CTTTGTATTTTGGAGGTGCTGCTGTTGGCATTTAGTTCAAGAGATTTCAGATTTCTG 706
Qy 41 LeuLeuGlyLysSerLeuAsp11eAlaProGlyAspMetLeuAspLysVal1AaArg 60
Db 707 CTTCTTGAAAGCTTTTGGACATAGCACAGGTGACATGCTTGAACAAGTGGCGAAGA 766
Qy 61 LeuSerLeu1eLysHisProGlyuCySerThrMetSerGlyLysVal1AaLeuHis 80
Db 767 CTTTCTTAATAAACAATCCAGAGTCTCCACATGATGATGGAAGCAATGAAACAT 826
Qy 81 LeuAlaLysGlnGlyAsnArgPheHisPheAsp11eLysProProLeuHisHisAlaLys 100
Db 827 TTGGCCAAACAGGAATAGATTTCATTTTGAATCAATCAATCCCTTGATCATGCTTAA 886
Qy 101 AsnCyAspPheSerPheThrGlyLeuGlnHisVal1ThrAspLys1eLleMetLysLys 120
Db 887 AATTGTGATTTTCTTTTACTGACTTCAACAGTTACTATTAATAAATAATGAAGACAG 946
Qy 121 GluLysGlnGlyLys1eLysGlyLysGln1eLeuSerSerAlaAlaAsp11eAla 140
Db 947 AAACAGAGAGAGATGATGAGAGGGGCAAAATCTGCTTCAGCAGCAGACATTTGCTGCC 1006
Qy 141 ThrValGlnHisThrMetAlaCyHisLeuValLysArgThrHisArgAla1eLeuPhe 160
Db 1007 ACAGTACAGCACATGATGCTCATCTTGTGAAAAGAACACATCCGGCTATTCGT 1066
Qy 161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaVal1LeuVal1AaSerGlyLysVal 180
Db 1067 TGTAAAGAGAGACTTGTACTTCAAAATTAATGCAATGCTGTTGATGCTGGTGGTCTG 1126
Qy 181 AlaSerAsnPheTyrlleArgArgAlaLeuGln1eLeuThrAsnAla1eThrGlnCysThr 200
Db 1127 GCAAGTAACTTCTATATCCAGAGGCTCTGAAATTTTAAACAACGCAACAGTGCAT 1186
Qy 201 LeuLeuCyProProProArgLeuCyThrAspAsnGlyLysMetLleAla1eArgAsnGly 220
Db 1187 TTGTGTGTCTCTCCACATGATGCACTGATATGATGATGATGATGATGATGATGAT 1246
Qy 221 IleGluArgLeuArgAlaGlyLeuGlyLys1eLeuHisAsp11eGlnGlyLys1eArg 240
Db 1247 ATTGAAGACTAGTGTGCTGGCTTGGCATTTTACATGACATAGAAAGCATCCGCTATGAA 1306
Qy 241 ProLysCyProLeuGlyValAsp11eSerLysGlnVal1GlyLys1eLeuSerLysVal 260
Db 1307 CCAAAATGCTCTCTTGAGATGACATATCAAAAGAAAGTTGAGAGAGCTTCCATTAAGTA 1366
Qy 261 ProGlnLeuLysMetGlu1e 267
```

```
Db      1367  CCACAAATTAATAATGAGATA 1387
RESULT 15
US-10-094-749-400
; Sequence 400. Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOTYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: Patent Ver. 2.1
; SEQ ID NO 400
; LENGTH: 2208
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-094-749-400

Alignment Scores:
Pred. No.:      1,47e-145      Length:      2208
Score:          1204.00      Matches:      239
Percent Similarity: 89.51%      Conservative: 0
Best Local Similarity: 89.51%      Mismatches: 4
Query Match:      86.93%      Indels:      24
DB:              18          Gaps:      1

US-10-649-273-2_COPY_148_414 (1-267) x US-10-094-749-400 (1-2208)
QY      1  MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20
Db      785  ATGGAGGCTCATGCACTTAATTAGTTGACCAATAAGTAGAATTCCTTTTTCAGTT 844
QY      21  LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
Db      845  CTTTGTGATTTCTGGAGGCTGCTGTGTGGCATTAGTTCAAGGAGTTTCAGATTTTTCG 904
QY      41  LeuLeuGlyLysSerLeuAspIleAlaLaprogLysPheMetLeuAspLysValAlaArgArg 60
Db      905  CTTCTTGGAAGGTTCTTTGGACATAGCACCAAGTGACATGCTTGACAAAGTGCGAAGAA 964
QY      61  LeuSerLeuIleLysHisIleProGluCysSerThrMetSerGlyGlyLysAlaIleGluHis 80
Db      965  CTTCTTTTAATTAATAACATCCAGAGTGCTCCACCATGAGTGAGGAAAGCCATAGACAT 1024
QY      81  LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 100
Db      1025  TTGGCCAAACAAAGGAATAGATTTCATTTTGCATCAATCAAAACCTCCCTGCATCATCTAAA 1084
QY      101  AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120
```

```
Db      1085  AATGTGATTTCTCTTTACTGACTTCAACAGCTTACTGATTAATTAATTAATGAAG 1144
QY      121  GluLysGluGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaAla 140
Db      1145  GAAAAAGAGAGATTTTGAAGAGGGCAATCTCTCTTCAGACACAGACATTCGTGCC 1204
QY      141  ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160
Db      1205  ACAGTACACACACCAATGGCAGTGCATCTTGTAAGAAAGAACACATCGGGCTATTCTGTT 1264
QY      161  CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyVal 180
Db      1265  TGTAAAGCAGAGAGACTTGTACCTCAAAATATAGCAGTACTGGTGCATCTGGTGATGTC 1324
QY      181  AlaSerAsnPheThrIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThr 200
Db      1325  GCAAGTAACTTCTGTATCCGCAAGCTCTGGAAATTTTAACAAACCGACACAGTGCACT 1384
QY      201  LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaThrAsnGly 220
Db      1385  TTGTTGTGTCCTCTCCAGACTATGCATGATTAATGGCATTATGATTGCA----- 1435
QY      221  IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyArgIu 240
Db      1435  ----- 1435
QY      241  ProLysCysProLeuGlyValAspIleSerLysGluValGlyGluAlaSerIleLysVal 260
Db      1436  ---TGATGTCCTCTTGAGTAGACATATCAAAAGATTGGAGAACCTTCATTAAGTA 1492
QY      261  ProGlnLeuLysMetGluIle 267
Db      1493  CCACAAATTAATAATGAGATA 1513
```

Search completed: November 11, 2005, 03:16:33
Job time : 896.614 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: November 10, 2005, 16:54:39 ; Search time 189.512 Seconds
(without alignments)
2305.321 Million cell updates/sec

Title: US-10-649-273-2_COPY_148_414

Perfect score: 1385
Sequence: 1 MEAHALTRLNKVEPPPLV.....DISKREVGASIKVQLKMEI 267

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n.model -DEV=x1p
-Q=/cgn2_1/USFTO.spool_p/US10649273/runat_02112005_091339_15581/app.query.fasta_1.1429
-DB=Issued_Patents_NA -QPM=fastap -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10649273 @CGN 1.1 177 @unat 02112005_091339_15581 -NCPU=6 -ICPU=3
-NO MAP -LAREOUTRY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -MAIN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6
-Fgapext=7 -Ygapop=10 -Ygapext=0.5 -Delop=6 -Delext=7

Database :

Issued Patents NA:*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PTUS.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1385	100.0	1416	4	US-09-774-528-177
2	1385	100.0	1526	4	US-10-067-443-23
3	1385	100.0	2197	4	US-10-067-443-1
4	1358.5	98.1	1387	4	US-10-067-443-21
5	824.5	59.5	14364	4	US-10-067-443-20
6	308.5	22.3	94750	4	US-09-596-002-38
7	307	22.0	1053	4	US-09-540-236-806
8	291.5	21.0	1059	4	US-09-252-991A-884
9	291.5	21.0	1266	4	US-09-252-991A-801
10	289.5	20.9	1830121	4	US-09-557-884-1
11	289.5	20.9	1830121	4	US-09-643-990A-1
12	282	20.4	996	4	US-09-902-540-6612

C	13	282	20.4	2582	4	US-09-902-540-503	Sequence 503, App
	14	280.5	19.3	1074	4	US-09-543-681A-2341	Sequence 2341, Ap
	15	270.5	20.5	1315	1	US-08-087-797-1	Sequence 1, Appli
	16	268.5	19.4	1032	4	US-09-489-039A-2050	Sequence 2050, Ap
	17	261	18.8	1092	4	US-09-107-532A-2955	Sequence 2955, Ap
	18	254	18.3	1008	3	US-08-987-121A-5	Sequence 5, Appli
	19	254	18.3	1011	3	US-08-987-121A-3	Sequence 3, Appli
	20	250	18.1	1006	4	US-09-536-784-51	Sequence 51, Appl
	21	250	18.1	1064	4	US-09-536-784-51	Sequence 214, App
	22	250	18.1	10974	3	US-08-961-527-214	Sequence 1, Appli
	23	248	17.9	1011	3	US-09-066-512-1	Sequence 2196, Ap
	24	246	17.8	1011	4	US-09-583-110-2196	Sequence 1618, Ap
	25	241	17.4	1011	4	US-09-107-433-1618	Sequence 1, Appli
	26	236.5	17.1	640681	4	US-09-790-988-1	Sequence 727, App
	27	236	17.0	822	4	US-09-710-279-727	Sequence 1551, Ap
	28	236	17.0	1101	4	US-09-134-000C-1551	Sequence 3985, Ap
	29	236	17.0	3993	4	US-09-134-000C-1551	Sequence 1072, Ap
	30	233	16.8	1107	3	US-09-134-001C-1072	Sequence 1, Appli
	31	220	15.9	1026	3	US-09-149-624-1	Sequence 794, App
	32	218	15.7	3064	3	US-09-221-017B-794	Sequence 1, Appli
	33	218	15.7	1230025	4	US-09-198-452A-1	Sequence 102, App
	34	218	15.7	1230230	4	US-09-438-185A-1	Sequence 102, App
	35	216	15.6	15249	4	US-08-956-171E-102	Sequence 2, Appli
	36	216	15.6	15249	4	US-08-781-986A-102	Sequence 147, App
	37	215.5	15.6	4403765	3	US-09-103-840A-2	Sequence 1, Appli
	38	215.5	15.6	4411529	3	US-09-103-840A-1	Sequence 1, Appli
	39	211.5	15.3	1155	4	US-09-602-777A-117	Sequence 1, Appli
	40	207.5	15.0	1664976	4	US-08-916-421B-1	Sequence 6, Appli
	41	207.5	15.0	1664976	4	US-09-692-570-1	Sequence 131, App
	42	192.5	13.9	1663	4	US-09-620-312D-6	Sequence 261, App
	43	185.5	13.4	42325	4	US-08-311-731A-121	Sequence 19, Appl
	44	181.5	13.1	432	4	US-09-328-352-261	
	45	169	12.2	876	4	US-09-724-623-19	

ALIGNMENTS

RESULT 1
US-09-774-528-177

Sequence 177, Application US/09774528 P.D.
Patent No. 6743619 1-50N-2604

GENERAL INFORMATION:

APPLICANT: Tang, Y. Tom
APPLICANT: Zhou, Ping
APPLICANT: Goodrich, Ryle
APPLICANT: Liu, Chenghua
APPLICANT: Asundi, Vinod
APPLICANT: Ren, Feiyan
APPLICANT: Zhang, Jie
APPLICANT: Zhao, Qing A.
APPLICANT: Yang, Yonghong
APPLICANT: Xue, Aidong J.
APPLICANT: Wehrman, Tom
APPLICANT: Wang, Jian-Rui
APPLICANT: Wang, Dunrui
APPLICANT: Dymnac, Radoje T.
TITLE OF INVENTION: No. 6743619el Nucleic Acids and
FILE REFERENCE: Polypeptides
CURRENT APPLICATION NUMBER: US/09/774,528
CURRENT FILING DATE: 2001-01-30
NUMBER OF SEQ ID NOS: 441
SOFTWARE: pc_fl_genes Version 2.0
SEQ ID NO 177
LENGTH: 1416
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE: CDS
NAME/KEY: (205) ..(1305)
LOCATION: (205) ..(1305)
US-09-774-528-177

Alignment Scores:

Pred. No.: 2,516-177 Length: 1416
Score: 1385.00 Matches: 267
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0

US-10-649-273-2_COPY_148_414 (1-267) x US-09-774-528-177 (1-1416)

```
QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20
DB 502 ATGAGGCTCAATGACCTTACTATTAGGTGACCAATTAAGTAATTCCTTTTATGTT 561
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
DB 562 CTTTGATTTCTGAGGCTGACCTGCTGTGGCATTAGTTCAAGAGGATTTCTG 621
QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60
DB 622 CTTCTGGAAGCTCTTGACATAGCACAGGTGACATGCTTGACAGAGTGCAAGAAGA 681
QY 61 LeuSerLeuIleLysHisProGluCysSerThrMetSerGlyLysValAlaIleGluHis 80
DB 682 CTTTCTTAATTAACATCCAGAGTCTCCACCAATGAGTGGGAAAGCCATTAGACAT 741
QY 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 100
DB 742 TTGGCCAAACAGAAATAGATTTCATTTTGACATCAAACTCCCTTGATCATGCTAAA 801
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120
DB 802 AATTGTGATTTTCTTTACTGACCTTCAACGCTTACTGATTAATATATATGAAGAAG 861
QY 121 GluLysGluGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaAla 140
DB 862 GAAAGAGGAAAGATTTGAGAGGGGCAAAATCTGTCTTGACAGACGACATTCCTGTC 921
QY 141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160
DB 922 ACAATGACGACACACATGACATGATCTTTGAAAAGAACACATGGGCTATTCGTTT 981
QY 161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyVal 180
DB 982 TGTAGACGAGAGACTTCTTACCTCAAAATATAGCATGCTGTTGCATCTGTGTGTC 1041
QY 181 AlaSerAsnPheTyrlleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThr 200
DB 1042 GCAAGTAACTTCTATATCCGACAGACTCTGGAAATTTTAAACAACGCAACAGGCACT 1101
QY 201 LeuLeuCysProProProArgLeuLeuCysThrAspAsnGlyIleMetIleAlaThrAsnGly 220
DB 1102 TTGTGTGTCTCTCCCTCCAGACTATGACCTGATTAATGGATTATGATTCAGTAATGCT 1161
QY 221 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrglu 240
DB 1162 ATTGAAGACTACGCTGCTGGCTTGGCATTTTACATGACATGAAGAGCATCCGCTATGA 1221
QY 241 ProLysCysProLeuGlyValAspIleSerLysGluValGlyGlnAlaSerIleLysVal 260
DB 1222 CCAAAATGCTCTCTTGAGTAGACATATCAAAAGAAAGTTGAGAGCTTCATTAAGA 1281
QY 261 ProGluLeuLysMetGluIle 267
DB 1282 CCACAAATTAATAATGAGATA 1302
```

RESULT 2
US-10-067-443-23

; Sequence 23, Application US/10067443
; Patent No. 6642041

; GENERAL INFORMATION:

; APPLICANT: Bristol-Myers Squibb Company

; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN
; TITLE OF INVENTION: SPINAL CORD, ME-1

```
/ FILE REFERENCE: D0073 NP  
/ CURRENT APPLICATION NUMBER: US/10/067,443  
/ CURRENT FILING DATE: 2002-02-05  
/ PRIOR APPLICATION NUMBER: US 60/266,518  
/ PRIOR FILING DATE: 2001-02-05  
/ PRIOR APPLICATION NUMBER: US 60/282,814  
/ NUMBER OF SEQ ID NOS: 71  
/ SOFTWARE: PatentIn version 3.0  
/ SEQ ID NO 23  
/ LENGTH: 1526  
/ TYPE: DNA  
/ ORGANISM: homo sapiens  
US-10-067-443-23
```

Alignment Scores:
Pred. No.: 2,856-177 Length: 1526
Score: 1385.00 Matches: 267
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0

US-10-649-273-2_COPY_148_414 (1-267) x US-10-067-443-23 (1-1526)

```
QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20
DB 1 ATGAGGCTCAATGACCTTACTATTAGGTGACCAATTAAGTAATTCCTTTTATGTT 60
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
DB 61 CTTTGATTTCTGAGGCTGACCTGCTGTGGCATTAGTTCAAGAGGATTTCTG 120
QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60
DB 121 CTTCTGGAAGCTCTTGACATAGCACAGGTGACATGCTTGACAGAGTGCAAGAAGA 180
QY 141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160
DB 181 CTTTCTTAATTAACATCCAGAGTCTCCACCAATGAGTGGGAAAGCCATTAGACAT 240
QY 161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyVal 180
DB 241 TGTAGACGAGAGACTTGTACTCTCAAAATTAAGCATGCTGTGTGATCTGTGTGTC 540
QY 181 AlaSerAsnPheTyrlleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThr 200
DB 541 GCAAGTAACTTCTATATCCGACAGACTCTGGAAATTTTAAACAACGCAACAGGCACT 600
QY 201 LeuLeuCysProProProArgLeuLeuCysThrAspAsnGlyIleMetIleAlaThrAsnGly 220
DB 601 TTGTGTGTCTCTCCCTCCAGACTATGACCTGATTAATGGATTATGATTCAGTAATGCT 660
QY 221 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrglu 240
DB 661 ATTGAAGACTACGCTGCTGGCTTGGCATTTTACATGACATGAAGAGCATCCGCTATGA 720
```

```
QY 241 ProlyscysProlenglyValAspIleSerlysglyValglyValAlaSerIlelyVal 260
DB 721 CCAAAAGTCCCTTGAGTACATATCAAAAGAGTTGAGAGCTTCCATAAAGTA 780
QY 261 ProGlnLeuLyMetGluIle 267
DB 781 CCACATTTAAATAAGAGATA 801

RESULT 3
US-10-067-443-1
; Sequence 1, Application US/10067443
; Patent No. 6642041
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 2197
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (231)..(1472)
US-10-067-443-1

Alignment Scores:
Pred. No.: 5,23e-177 Length: 2197
Score: 1385.00 Matches: 267
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0

US-10-649-273-2_COPY_148_414 (1-267) x US-10-067-443-1 (1-2197)
QY 1 MetGluAlaHisAlaIleuThrIleArgLeuThrAsnLyValGluPheProPheLeuVal 20
DB 672 ATGAGAGGCTCAAGCACTTACTATTAGGTGACCAATTAAGTAAGTAATTCCTTTTATGTT 731
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
DB 732 CTTTGTGATTTCTGAGAGTCACTGCTCTGTTGGCATTAGTTCAAGAGATTTCAGATTTTCTG 791
QY 41 LeuLeuGlyLySerLeuAspIleAlaProGlyAspMetLeuAspLyValAlaArgArg 60
DB 792 CTTCTTGGAAGGCTTTTGACATAGCACAGGTGACATGCTTGACAGAGGTGGCAAGAGA 851
QY 61 LeuSerLeuIleLyHisProGlyCysSerThrMetSerGlyGlyValAlaIleGluHis 80
DB 852 CTTTCTTTAAATTAACATCCAGAGGTCTCCACATGATGGTGGGAAAGCCCTACAGACAT 911
QY 81 LeuAlaLySGlnGlyAsnArgPheHisAspIleLySerProPheLeuHisAlaLyS 100
DB 912 TTGGCCAAACAGGAATAGATTTCATTTTGAACATCAAACTCCCTTGATCATCTAA 971
QY 101 AsnCyAspPheSerPheThrGlyLeuGlnHisValThrAspLySileIleMetLySlyS 120
DB 972 AATTGTGATTTTCTTTTACTGAGACTTCACACGTTACTGATTAATAATATGAAAAAG 1031
QY 121 GlnLySGlnGlyIleGlnLySGlyGlnIleLeuSerSerAlaAlaAspIleAlaAla 140
DB 1032 GAAAAAGAGGAAGATATGAGAGAGGGAATCCGTCTTCAGCAGACATTCCTGCC 1091
QY 141 ThrValGlnHisThrMetAlaCysHisLeuValLySArgThrHisArgAlaIleLeuPhe 160
```

```
DB 1092 ACAGTACAGCACACAAATGGCATGCTTGTGAAAAAACAATGGGCTATTCGTTT 1151
QY 161 CysLySGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyValAl 180
DB 1152 TGTAAAGCAGAGAGACTTGTACTTCAAAATATGCACTGCTGTTGCATCTGATGATGTC 1211
QY 181 AlaSerAsnPheThrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThr 200
DB 1212 GCAAGTAACTTCTATATTCGACAGAGCTTGAAATTTTAAACAAACCAACACAGTCACT 1271
QY 201 LeuLeuCySPProProPheArgLeuCyThrAspAsnGlyIleMetIleAlaTrpAsnGly 220
DB 1272 TTGTTGTGCTCCCTCCACAGCTATGCATGATTAATGGCATTATGATTCATGGAATGGT 1331
QY 221 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlnGlyIleArgTyrglu 240
DB 1332 ATTTGAAGACTACAGCTGCTGGCTTGGCATTATTAACATGACATAGAGGCATCCGCTATGAA 1391
QY 241 ProLyScysProlenglyValAspIleSerlysglyValglyValAlaSerIlelyVal 260
DB 1392 CCAAAATGCTCCTTGAGTACATATCAAAAGAGTTGAGAGACTTCCATTAAGAATA 1451
QY 261 ProGlnLeuLyMetGluIle 267
DB 1452 CCACATTTAAATAAGAGATA 1472

RESULT 4
US-10-067-443-21
; Sequence 21, Application US/10067443
; Patent No. 6642041
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 21
; LENGTH: 1387
; TYPE: DNA
; ORGANISM: homo sapiens
US-10-067-443-21

Alignment Scores:
Pred. No.: 9.58e-174 Length: 1387
Score: 1358.50 Matches: 266
Percent Similarity: 91.10% Conservative: 1
Best Local Similarity: 98.09% Mismatches: 1
Query Match: 98.09% Indels: 25
DB: 4 Gaps: 1

US-10-649-273-2_COPY_148_414 (1-267) x US-10-067-443-21 (1-1387)
QY 1 MetGluAlaHisAlaIleuThrIleArgLeuThrAsnLyValGluPheProPheLeuVal 20
DB 465 ATGAGAGGCTCAAGCACTTACTATTAGGTGACCAATTAAGTAAGTAATTCCTTTTATGTT 524
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
DB 525 CTTTGTGATTTCTGAGAGTCACTGCTCTGTTGGCATTAGTTCAAGAGATTTCAGATTTTCTG 584
QY 41 LeuLeuGlyLySerLeuAspIleAlaProGlyAspMetLeuAspLyValAlaArgArg 60
DB 585 CTTCTTGGAAGGCTTTTGACATAGCACAGGTGACATGCTTGACAGAGGTGGCAAGAGA 644
QY 61 LeuSerLeuIleLyHisProGlyCysSerThrMetSerGlyGlyValAlaIleGluHis 80
```

|||||
Db 645 CTTTCTTAAATAAACATCCAGAGCTCCACCATGAGTGGGAGAAAGCCATTAGACAT 704
Qy 81 LeuAlaIysGlnGlyAsnArgPheHisPheAspIleuYsProProLeuHisIleAlaIys 100
Db 705 TTGGCCAAACAGAAATAGATTTCATTTTGCATCAAACTCCCTTGCAATATCTTAA 764
Qy 101 AsnGlyAspPheSerPheThrGlyLeuGlnHisValIthrAspIysIleIleMetLysIys 120
Db 765 AATTGTGATTTTCTTTACTGACTTCAACAGCTTACTGATTAATAATATATGAAAAAG 824
Qy 121 GlnIysGlnGlnGlyIle----- 126
Db 825 GAAAGAGAGAGATATTTCTAATTAGTAAAGTTGAAACAGATTAATTCCTGCAATG 884
Qy 127 -----GlnIysGlnGlnIleLeuSerSerIle 135
Db 885 TGCCATAAAATAGCTGCTCATTTTCTGCAAGTATGAGAAAGGCAAAATCTGCTTCAGCA 944
Qy 136 AlaAspIleAlaIleAlaIleValGlnHisIleMetAlaCysHisIleValIleArgThrHis 155
Db 945 GCAGACATTCCTGCGACAGTACAGACACAAATGSCATCTTGTGAAAAAGAACAT 1004
Qy 156 ArgAlaIleLeuPheCysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValIleVal 175
Db 1005 CGGCGTATTCCTTTTGTAAACAGAGACATTTGTTACTCAAAATTAAGCAGTACTGTT 1064
Qy 176 AlaSerGlyValAlaIleSerAspPheThrIleArgArgAlaIleGlnIleLeuThrAsn 195
Db 1065 GCATCTGTGTGTGTGCGAAGTACTTCTATATCCGCAAGCTCTGAAATTTTAAACAAC 1124
Qy 196 AlaThrGlnCysThrLeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMet 215
Db 1125 GCAACACAGTGCATTTTGTGTCTCTCCACACATGACATGATATATGCACTTATG 1184
Qy 216 IleAlaIleAsnGlnIleGlnArgLeuArgAlaGlnIleLeuHisAspIleGln 235
Db 1185 ATTCATGAGATGTGATGAAAGACTAGCTGTGCTGGGCACTTTTACATGACATAGAA 1244
Qy 236 GlnIleArgGlyArgProLeuCysProLeuGlnValAspIleSerLysGlnValGln 255
Db 1245 GGCATCCCGCTATGAAACCAAAATGCTCTTGGAGTACATATCAAAAGAAAGTTGGAGA 1304
Qy 256 AlaSerIleLysValProGlnLeuLysMetGluIle 267
Db 1305 GCTTCATTAAGTACCAATTAATAATGAGATA 1340
RESULT 5
US-10-067-443-20
; Sequence 20, Application US/10067443
; Patent No. 6642041
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN
; FILE REFERENCE: SPINAL CORD, MP-1
; CURRENT APPLICATION NUMBER: D0073 NP
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 20
; LENGTH: 14364
; TYPE: DNA
; ORGANISM: homo sapiens
US-10-067-443-20
Alignment Scores: 1.34e-99 Length: 14364
Pred. No.: 824.50 Matches: 240

Percent Similarity: 26.73% Conservative: 0
Best Local Similarity: 26.73% Mismatches: 2
Query Match: 59.53% Indels: 657
DB: 4 Gaps: 4
US-10-649-273-2_COPY_148_414 (1-267) x US-10-067-443-20 (1-14364)
Qy 1 MetGlnAlaHisAlaIleuThrIleArgLeuThrAsnLysValGlnPheProPheLeuVal 20
Db 10845 ATGAGAGCTCATGCACTTACTATTAAGTTGACCAATPAAGTAAATTCCTTTTAGTT 10904
Qy 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaIleValGlnGlyValSerAspPheLeu 40
Db 10905 CTTTGTATTTCTGAGAGTCACTGTCTGTGGCAATTAGTTCAGAGAGTTTCAGATTTTCTG 10964
Qy 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysVal----- 57
Db 10965 CTTCAATGAAAGCTTTTGGACATGACACAGGTGACATGCTTGACAAAGGT-AAATTAAAG 11023
Qy 57 ----- 57
Db 11024 TTAATTTTCATTCCTTTTGTATAGTTGTCATTTCACATPAAGTATGATGTGC 11083
Qy 57 ----- 57
Db 11084 TACCAACATTCACCTAAATATTTCTGAATTTATCTTAGTAAACGAAAAATTCACAT 11143
Qy 57 ----- 57
Db 11144 ATGTGAGAAAAAATGAAAAAGTATGACAAATTTATATTTCTTAGCCTTTCTTAATA 11203
Qy 57 ----- 57
Db 11204 AAATGTAAAGAGTTTCATATCTGTACATAAAGCTGAATATGTTGCAATACATTAAG 11263
Qy 57 ----- 57
Db 11264 TATTTTCCAAATATATGTATGTGAAAAAGAGTCTGTAACTAAATACATACGAAAAA 11323
Qy 57 ----- 57
Db 11324 GGTAAATTAAGAAATATATATATATATTAATTAATTAAGACATTAAGATGCAATGACAGAA 11383
Qy 57 ----- 57
Db 11384 TTAATTAACAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 11443
Qy 57 ----- 57
Db 11444 AATTAATCAAGAGCTACTGCAATATATATAGAAAAAACAACAACAACAACAACAACATG 11503
Qy 57 ----- 57
Db 11504 CTTCCACAGTAAATTAATTAAGAAATTAAGAAAGTCTTATTAATTAAGCTTCAATCATT 11563
Qy 57 ----- 57
Db 11564 AAGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 11623
Qy 57 ----- 57
Db 11624 GGGCAGTATGAGCTCTTATTTGTCGGGTCAATATAGGAACAGGTTGTCTGTACG 11683
Qy 57 ----- 57
Db 11684 TGAATATCACTATATAGTCTATATTTGCAAAAGTATAGCATGTTTATTCATGAGGGTT 11743
Qy 57 ----- 57
Db 11744 TTTTGTGTTGTTAGTATTTTCAATTTATTTCTTTGATCTTTGCTTTCACAGTATTT 11803
Qy 58 -----AlaArgArgLeuSerLeuIle 64
|||||

Db	11804	AATTTTAGCTCAATAAAAAATATGTTCTTTGATAGTAGGCAAGAAGACCTTCTTAATA	11865
Qy	65	LyshiProGluCySerSerThiMetSeSelGlyIlyseAlileGluIshleuAlalyGln	84
Db	11864	AAACATCCAGAGTGTCTCCACCATAGAGTGGGAAGCCATAGAACATTTGGCCAAACA	11923
Qy	85	GlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlalyAsnCyAspPhe	104
Db	11924	GGAAATAGATTCTATTGTGACATACACCTCTTGATCATGTCTAAAAATGTGATTT	11983
Qy	105	SerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGln	124
Db	11984	TCTTTTACTGGACTTCAACACGTTACTGATATAAATAATATGAAAAAGAGAA	12043
Qy	124	-----	124
Db	12044	GGTATATTTCTAATTAAGTAAAGTTGAACAGATAAATATTCCTGGATTGTCCCTAAAAATA	12103
Qy	125	-----GlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAl	139
Db	12104	GCTGCTCATTTCTTGCAAGTATTTGAGAGGGGCAATCTGTCTTGACGACAGACATGG	12165
Qy	139	alaIThrValGlnHisThrMetAlaCyHisLeuValLysArgThrHisArgAlaIleLe	159
Db	12164	TGCCACAGTACAGACACCAATGGATGTCACTCTGTGAAGAAGAACATCGGCTATTTCT	12223
Qy	159	uPheCyLeuLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeu-----	174
Db	12224	GTTTTGTAGACAGAGACTTGTATCTCAAAAATAAGCAGTACTGTATAGTTTATCTTC	12283
Qy	174	-----	174
Db	12284	ATTTTATAGTAAATAGTTACACTTTGCAATATGTTACTTTTCCCAAGACCTTGACCTTG	12343
Qy	174	-----	174
Db	12344	TGTTTAGATGAACAGATCTTTATAGCCTTATGCTAGCCCTGACAGATAGAAATTATGCAG	12403
Qy	174	-----	174
Db	12404	GATAGAAAGACTAAACAGCAATTTCTGTACTAGTTGTAGCTTATGGGACAGCTGTA	12463
Qy	174	-----	174
Db	12464	TAGCTTCTATGGACATAGTCTAATTTTGCATCTTCTGTGSAATTAAGAAGGGCTT	12523
Qy	174	-----	174
Db	12524	ACATTAAGAAGAAAGTAATGCAGTAACTGCTATCATATTTTGAAGAAAAAGTGCAATTT	12583
Qy	174	-----	174
Db	12584	CCTTCATCTTTGATGAATTCCTTTGTGTGTGTTTTTATATAGCAGCAAAATTT	12643
Qy	174	-----	174
Db	12644	AGCAGTGGAGTGTATTCCAACTTTCGTAGACATAATGTGATAAAGTTCTGATATATC	12703
Qy	174	-----	174
Db	12704	CACATATTTGACAGCCAAATTCCTTTAATGTGTAAAAAGCTTGACAAACATCCT	12763
Qy	174	-----	174
Db	12764	GTTTAACGTATCTTAACCTTATATTCATTTAAAAATATATAACTAAAGTGGAAATGTT	12823
Qy	174	-----	174
Db	12824	TAAATGTATTAATTCATAGATGAATTTTACATGATATCAAGAATATATTTTTCAGA	12883
Qy	174	-----	174
Db	12884	GTTATGTAGTAAAAAGCAAAAATATATAAAAATTTCAAGGCTTAATAATATAGTACTATATGA	12943

[illegible]


```

; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 884
; LENGTH: 1059
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-884

Alignment Scores:
Pred. No.: 3,67e-29      Length: 1059
Score: 291.50           Matches: 87
Percent Similarity: 50.41% Conservative: 37
Best Local Similarity: 35.37% Mismatches: 99
Query Match: 21.05%     Indels: 23
DB: 4                   Gaps: 9

US-10-649-273-2_COPY_148_414 (1-267) x US-09-252-991A-884 (1-1059)
QY      1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLys---ValGluPheProPheLeu 19
Db      711 ATGGAAGGGCACCTGCTGGCGCCGATGCTGGAAAGCAGCCACCGCGGTTCCCGTTCGTC 652
QY      20 ValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPhe 39
Db      651 GCCTTGCTGGTTTCCGGCGGTACACCCAGTTGGTGGCGGTGGAGGATGCGCGCGCTAC 592
QY      40 LeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspIlyValAlaArg 59
Db      591 CAGTTGCTGGCGAATCGGTGACGATGCGCGCGCGAAGCCTTGACAAAGCCGCAAG 532
QY      60 ArgLeuSerLeuIleLysHisProGluCysSerThrMetSerGlyGlyLysAlaIleGlu 79
Db      531 CTGATCGGCTG---GGCTATCCC-----GGTGTCTCGGAATTCGCC 493
QY      80 HisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAla 99
Db      492 CGCTGGCGAGCGCGGACACTCTGGCGGCTTCGTTCCCGGCGCGATGACCGATGCC 433
QY      100 LysAsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLys 119
Db      432 CCGCGCTGAGACTTCAGCTTCAGCGGCTCAAGACCTTACCTTAAC---ACCTGGCAG 376
QY      120 LysGluLysGluGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAla 139
Db      375 CGTTGCGTGGAGCGCGGACGACGACGACGAG-----ACCGCTGGCAGATCGCC 325
QY      140 AlaThrValGlnHisThrMetAlaCysHisIleValLysArgThrHisAlaAlaIleLeu 159
Db      324 CTGGCGTTCCAGACCGCGGATGTCAGACCTGCTGATCAAGTGCCTGCGCCCTTG--- 268
QY      160 PheCysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValIleValAlaSerGlyGly 179
Db      267 -----AAGCAGACCGGCTG-----AAGAAC-----CTGATATCGCGCGCGGT 229
QY      180 ValAlaSerAsnPheTyrlleArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCys 199
Db      228 GTACGCCCAACCAAGCGCTGCGCAGCGGCTGGAAAGATGCTCGCGGAATGAAGGGG 169
QY      200 ThrLeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaTrpAsn 219
Db      168 CAGGTGTTCTAAGCCCGCGCGCTTCTGACCGACCAATGCGCGCATATCGCTACGCC 109
QY      220 GlyIleLeuArgLeuArgAlaGlyIleuGlyIleLeuHisAspIleGluGlyIleArgTyx 239
Db      108 GGTGCGCAGCGCTGCTGCGCGC-----CAGCATAGAGCGCGCGCATACGCGCT 58
QY      240 GluProLysCysProLeu 245
Db      57 CAGCGCGCGCTGCGCGAGT 40
```

```

RESULT 9
US-09-252-991A-801
; Sequence 801, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 801
; LENGTH: 1206
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-801

Alignment Scores:
Pred. No.: 4.55e-29      Length: 1206
Score: 291.50           Matches: 87
Percent Similarity: 50.41% Conservative: 37
Best Local Similarity: 35.37% Mismatches: 99
Query Match: 21.05%     Indels: 23
DB: 4                   Gaps: 9

US-10-649-273-2_COPY_148_414 (1-267) x US-09-252-991A-801 (1-1206)
QY      1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLys---ValGluPheProPheLeu 19
Db      514 ATGGAAGGGCACCTGCTGGCGCCGATGCTGGAAAGCAGCCACCGCGGTTCCCGTTCGTC 573
QY      20 ValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPhe 39
Db      574 GCCTTGCTGGTTTCCGGCGGTACACCCAGTTGGTGGCGGTGGAGGATGCGCGCGCTAC 633
QY      40 LeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspIlyValAlaArg 59
Db      634 CAGTTGCTGGCGAATCGGTGACGATGCGCGCGGGAAGCCTTGACAAAGCCGCAAG 693
QY      60 ArgLeuSerLeuIleLysHisProGluCysSerThrMetSerGlyGlyLysAlaIleGlu 79
Db      694 CTGATCGGCTG---GGCTATCCC-----GGTGTCTCGGAATTCGCC 732
QY      80 HisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAla 99
Db      733 CGCTGGCGAGCGCGGACACTCTGGCGGCTTCGTTCCCGGCGCGATGACCGATGCC 792
QY      100 LysAsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLys 119
Db      793 CCGCGCTGAGACTTCAGCTTCAGCGGCTCAAGACCTTACCTTAAC---ACCTGGCAG 849
QY      120 LysGluLysGluGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAla 139
Db      850 CTTGCGTGGAGCGCGGACGACGACGACGAG-----ACCGCTGGCAGATCGCC 900
QY      140 AlaThrValGlnHisThrMetAlaCysHisIleValLysArgThrHisAlaAlaIleLeu 159
Db      901 CTGGCGTTCCAGACCGCGGATGTCAGACCTGCTGATCAAGTGCCTGCGCCCTTG--- 957
QY      160 PheCysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValIleuValAlaSerGlyGly 179
Db      958 -----AAGCAGACCGGCTG-----AAGAAC-----CTGATATCGCGCGCGGT 996
QY      180 ValAlaSerAsnPheTyrlleArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCys 199
Db      997 GTACGCCCAACCAAGCGCTGCGCAGCGGCTGGAAAGATGCTCGCGGAATGAAGGGG 1056
```

Qy 200 ThrlleuLeuCySProProAryGleuCyThraPaanglylleMetllea1aTtpasn 219
Db 1057 CAGGTGTTCTTAGCCCGCCGCTTCTGCACCGCAATGGCGCATGATCGCTTACGCC 1116
Qy 220 G1y1leG1uAryGAlaG1y1leuG1y1leuH1eap1leG1uG1y1leArg1y 239
Db 1117 GGGTCCAGCCGCTGCTGCGCGC-----CAGCATGACGGCCGCGCATGCGCTC 1167
Qy 240 GluProlySycSProleu 245
Db 1168 CAGCCGCGCTGCGCGCATG 1185
RESULT 10
US-09-557-884-1/c
; Sequence 1, Application US/09557884
; Patent No. 6506581
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: The Nucleotide sequence of the Haemophilus influenzae Rd Genome, Fragments Thereof, and Uses Thereof
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3 1/2 inch diskette
; COMPUTER: Dell Pentium
; OPERATING SYSTEM: MS DOS v6.22
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/557,884
; FILING DATE: 25-Apr-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/476,102
; FILING DATE: JUN-5-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; TELECOMUNICATION INFORMATION:
; TELEPHONE: 301-309-8504
; TELEFAX: 301-309-8439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1830121 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1
Alignment Scores:
Pred. No.: 1.68e-23 Length: 1830121
Score: 289.50 Matches: 82
Percent Similarity: 48.70% Conservative: 30
Best Local Similarity: 35.65% Mismatches: 95
Query Match: 20.90% Indels: 23
Gaps: 7
US-10-649-273-2_COPY_148_414 (1-267) x US-09-557-884-1 (1-1830121)
Qy 1 MetG1uA1aH1eA1aLeuThrl1eArg1eU---ThraS1yVa1G1u1heProPhe1eU 19.
Db 552881 ATGAAAGGCAATTATTCGCGCAATGATGACAAATTCACCGCATCTTTTGT 552882
Qy 20 Val1euLeu1leSeRg1yG1yH1eCyS1euLeuA1aLeuVa1G1uG1yVa1SeR1aPhe 39

Db 552821 GCTCTGTTGCTATCGGCGCCACACTCAATTAGTCGCTGATGCTAGAGAAATAT 552762
Qy 40 Leu1euLeuG1yLySeR1euA1aProG1yA1aPhe1eA1aArg 59
Db 552761 GAAGTATAGAGAAATCTATTATGATGCTGCGGAAGCCCTTGATTAACGAGAAA 552702
Qy 60 Arg1euSeR1eU1leY1eH1eS1eProG1yCySeR1eM1eSeRg1yG1ySa1a1eG1u 79
Db 552701 TTACTTGGACTA---GATTATCCA-----GTTGCGCGGACCTTTCT 552663
Qy 80 H1e1eA1aLyG1uG1yA1aRg1eH1eS1ePheA1a1eLyS1eProPhe1eH1eS1eA1a 99
Db 552662 CATTAGCGGAAAAGTACGCCAAATCGTTTACATTCCTCAATGACAGATCGT 552603
Qy 100 LyS1eA1aCyS1ePheSeR1eH1eG1yG1uH1eS1eVal1ThraS1yS1e1leMet1yS 119
Db 552602 GCAAGCCTGATTTTATGTTTCTTCTGTTTAAACCTTGCCGCAATACATTATTCAA 552543
Qy 120 LyS1eG1yS1eG1uG1y-----1leG1yS1eG1yG1u1leuSeR1eA1aA1aP 137
Db 552542 GCAATTAATAAAGAGGCGCACTGATAGAGCAA-----ACTAAAGCAGAT 552498
Qy 138 1leA1a1aThVa1G1uH1eS1eThraA1aCyS1e1eUVal1yS1eRg1eH1eS1eA1a 157
Db 552497 ATTCCTTATGCTTCCAGATCGGCTGATCTTCTGCC----- 552456
Qy 158 1leuPheCyS1eG1uA1aRg1eA1aPhe1eUProG1uA1aS1a1aVal1euVa1a1aSer 177
Db 552455 ---ATTAAATGTAG---CGTGATTTGAAAGAAACAGCGCTATAAAGTTTAGATTCG 552402
Qy 178 G1yG1yVa1a1aSeR1aPhe1y1leA1aRg1a1aLeuG1u1leuThraS1a1aThr 197
Db 552401 GGAGGCGTGAAGCAATTAATAAAGCTCGAGAAACGCTTGCGCATTAATGCAAAATTTA 552342
Qy 198 G1uCyS1eThrlleuCySProProAryGleuCyThraPaanglylleMetllea1a 217
Db 552341 GGTGGGAGTGTATTATCTCAACTCAATTTGTAAGATATATGTCGATGATTCCT 552282
Qy 218 TTPaang1y1leG1uAryGAlaG1y 227
Db 552281 TACACAGGTTTTTTACGTTTAAACAGGT 552252
RESULT 11
US-09-643-990A-1/c
; Sequence 1, Application US/09643990A
; Patent No. 6528289
; GENERAL INFORMATION:
; APPLICANT: Robert D. Fleischmann
; Mark D. Adams
; Owen White
; Hamilton O. Smith
; J. Craig Venter
; TITLE OF INVENTION: The Nucleotide sequence of the Haemophilus influenzae Rd Genome, Fragments Thereof, and Uses Thereof
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville,
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3 1/2 inch diskette
; COMPUTER: Dell Pentium
; OPERATING SYSTEM: MS DOS v6.22
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/643,990A
; FILING DATE: 23-Aug-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:

```

      APPLICATION NUMBER: 08/487,429
      FILING DATE: 1995-06-07
      APPLICATION NUMBER: 08/426,787
      FILING DATE: 1995-04-21
      ATTORNEY/AGENT INFORMATION:
      NAME: Kenley K. Hoover
      REGISTRATION NUMBER: 40,302
      REFERENCE/DOCKET NUMBER: PB186PIC1
      TELECOMMUNICATION INFORMATION:
      TELEPHONE: 301-610-5790
      TELEFAX: 310-309-8439
      INFORMATION FOR SEQ ID NO: 1:
      SEQUENCE CHARACTERISTICS:
      LENGTH: 1830121 base pairs
      TYPE: nucleic acid
      STRANDEDNESS: double
      TOPOLOGY: linear
      SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1

Alignment Scores:
Pred. No.: 1,68e-23 Length: 1830121
Score: 289.50 Matches: 82
Percent Similarity: 48.70% Conservative: 30
Best Local Similarity: 35.65% Mismatches: 95
Query Match: 20.90% Indels: 23
DB: 4 Gaps: 7

US-10-649-273-2_COPY_148_414 (1-267) x US-09-643-990A-1 (1-1830121)
QY 1 MetGluAhiAalaleuThrlleArgleu--ThAsnlySvalGluPheProPheleu 19
Db 552881 ATGAGAGGCAATTCCTGCGCAATGCTGTATGACAAATTCACCGACTTCTCTTTGTT 552822
QY 20 ValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspHe 39
Db 552821 GCTCTGTGGTATCGGCGGCGCACCTCAATTAGCGCTGTGAGTGATGAGAAATAT 552762
QY 40 LeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArg 59
Db 552761 GAAGATAGAGAAATCTATTGATGATGCTGCTGCGAAGCCTTTGATTAACAGCAAAA 552702
QY 60 ArgLeuSerLeuIleLysHisProGlyCysSerThrmSerGlyGlyLysAlaIleGlu 79
Db 552701 TTACTTGACATG---GATTATCCA-----GTTGGCGCGGCACTTTCT 552663
QY 80 HisLeuAlaLysGlnGlyAsnArgPheHisAspAspIleLysProProLeuHisAla 99
Db 552662 CGTTTAGCGGAAAAAGGTACCGCAAAATCGTTTCACATTTCCAGTTCATGACAGATCGT 552603
QY 100 LysAsnCyAspPheSerPheThrclyLeuGlnHisValThraSplyleIleMetLys 119
Db 552602 GCAGCGCTTGATTTAGTTTCTTCTGTTTAAACATTTGCCGCAAAATACATTAATCA 552543
QY 120 LysGluLysGlnGluGly-----IleGluLysGlnIleLeuSerAlaIlaAsp 137
Db 552542 GCATTAATAAAACGAGGCGCAACTGATAGCA-----ACTAAAGCAAGT 552498
QY 138 IleAlaAlaThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAla 157
Db 552497 ATTCCTTAATGCTTCCAGAGACCGGTGATGATCTTTGCC----- 552456
QY 158 IleLeuPheCyLysGlnArgAspLeuProGlnAsnAsnAlaValLeuValAlaSer 177
Db 552455 ---ATTAAATGTAG---CGTGCATTGAAGAAACAGCGCTATAAACGTTTATGATTGCG 552402
QY 178 GlyValValAlaSerAspPheThrIleArgArgAlaLeuGlnIleLeuThrSrnAlaThr 197
Db 552401 GGAAGGAGGAGCGCAATAAAACCTCGAGAAACGCTTGCGCACTTAATGCAAAATTTA 552342
QY 198 GlnCyThrLeuLeuCyProProArgLeuCyThrAspAsnGlyIleMetIleAla 217
Db 552341 GGTGGCGAAGTGTATTATCTCAACCTCAATTTTGTATGATTAATGCGTGGATGTTGCT 552282
```

```

QY 218 TrpAsnGlyIleGlnArgLeuArgAlaGly 227
Db 552281 TACACAGTGTTTTACGTTTAAACAGGT 552252

RESULT 12
US-09-902-540-6612
; Sequence 6612, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; PRIORITY FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 6612
; LENGTH: 996
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-6612

Alignment Scores:
Pred. No.: 6.44e-28 Length: 996
Score: 282.00 Matches: 86
Percent Similarity: 48.93% Conservative: 28
Best Local Similarity: 36.91% Mismatches: 87
Query Match: 20.36% Indels: 32
DB: 4 Gaps: 9

US-10-649-273-2_COPY_148_414 (1-267) x US-09-902-540-6612 (1-996)
QY 1 MetGluAhiAalaleuThrlleArgleuThraSrnlyVal---GluPheProPheleu 19
Db 334 CTGAGAGGCGACCTGCTGCGCATCGGCTGTGAGAGTGGCGCGGAGCCGCTTCTT 393
QY 20 ValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspHe 39
Db 394 GGGCTGCTGTTTCCGCGGCGGACACAGCAGCTCTACAGAGTGCAGGCTTACCGCGAGTAC 453
QY 40 LeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArg 59
Db 454 CGGCTGCTGCGGCGAGCAGCGCGAGCAGCGCGCGGAGGAGTATGACAAAGCCGCTCC 513
QY 60 ArgLeuSerLeuIleLysHisProGlyCysSerThrmSerGlyGlyLysAlaIleGlu 79
Db 514 ATCTCGCGCTG---CCGTATCCG-----GTTGGAGGCCCATGAGC 552
QY 80 HisLeuAlaLysGlnGlyAsn-----ArgPheHisAspIleLysProPro 95
Db 553 CAGTTGCGCCACAGGAGGAAACCGAGGACCATCCGCTT-----CCGCC 597
QY 96 LeuHisAlaLysAsnCyAspPheSerPheThrclyLeuGlnHisValThraSplyLys 115
Db 598 GCGTGCAGGAGCAAACTTGCAGCGTCTTCTCGGAGTGAAG-----ACG 645
QY 116 IleIleMetLysLysGlnGluGlnGlyIleGlnLysGlnIleLeuSerAlaIlaAsp 135
Db 646 GCGGCTGCGCACACCTGAGAAAGCAGCGTGCAGCGGAGGAGCGCGCTG----- 696
QY 136 AlaAspIleAlaIleThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHis 155
Db 697 GCGAATTTGCGCGCTTCTTCCAGAGGCGCTGCGGAGAGCTGTGTGAGAAAG----- 750
QY 156 ArgAlaIleLeuPheCyLysGlnArgAspLeuProGlnAsnAsnAlaValLeuVal 175
Db 751 -----CTGTGGCGCGCGCGCGCGCTTG-----GCCACAAGAGTGTGGTG 792
```



```

Qy 120 LysGluLysGluGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAla 139
Db 724 AACCATGATTTCAGAG-----CAAACTCGACGACATATTCGG 759
Qy 140 AlathrValGlnHisThrMetAlaCysHisIleuValLysArgThrHisArgAlaIleLeu 159
Db 760 CGTCCTTTGAAGATCCCGTGTAGATCTTTGGCAATAAATGTCGTGACCA----- 813
Qy 160 PheCysLysGlnArgAspLeuProGlnAsnAsn---AlaValIleuValAlaSerGly 178
Db 814 -----TTAGACCAACAGCCTTTAAACGCTTGTGATGCTGG 852
Qy 179 GlyValAlaSerAspPheTyrIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGln 198
Db 853 GGCCCTTAAGTGTAAACCGTACTTTACGCCCAAAATGGCGATATATGAAACAACTCGGA 912
Qy 199 CysThrIleuLeuCysProProArgLeuCysThrAspAsnGlyIleMetIleAlaTrp 218
Db 913 GGGGAAAGTTTATGCTGCGCCGTAGATTATGTACCAATATGTCCTGATTCCTTGG 972
Qy 219 AsnGlyIleGluArgLeuArgAlaGlyLeu---GlyIleLeuHisAspIleGluGlyIle 237
Db 973 GCGGGAGATGCCGTTTAAAGTGTACCGAAGGGGCTTTA-----GGGGTG 1020
Qy 238 ArgTyrGluProLysCysProLeu 245
Db 1021 ACAGTGAGACCAAGTGGCTTTA 1044

```

```

RESULT 15
US-08-087-797-1
; Sequence 1, Application US/08087797
; Patent No. 5543312
; GENERAL INFORMATION:
; APPLICANT: Mellors, Alan
; APPLICANT: Lo, Reggie Y.C.
; APPLICANT: Abdullah, Khalid M.
; TITLE OF INVENTION: Pasteurella Haemolytica
; TITLE OF INVENTION: Glycoprotease
; TITLE OF INVENTION: Gene and the Purified Enzyme
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bell, Seltzer, Park & Gibson, P.A.
; STREET: 1211 East Morehead Street,
; CITY: Charlotte
; STATE: No. 5543312th Carolina
; COUNTRY: United States
; ZIP: 28234
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/087,797
; FILING DATE: 14-JUL-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Layton, Jr., Samuel G.
; REGISTRATION NUMBER: 22807
; REFERENCE/DOCKET NUMBER: 3374-80
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 704 377 1561
; TELEFAX: 704 334 2014
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1315 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; FEATURE:

```

```

; NAME/KEY: CDS
; LOCATION: 141..1115
; US-08-087-797-1
Alignment Scores:
Pred. No.: 3,72e-26 Length: 1315
Score: 270.50 Matches: 79
Percent Similarity: 46.46 Conservative: 26
Best Local Similarity: 34.96 Mismatches: 102
Query Match: 19.53 Indels: 19
DB: 1 Gaps: 6

```

US-10-649-273-2_COPY_148_414 (1-267) x US-08-087-797-1 (1-1315)

```

Qy 1 MetGluAlaHisAlaLeuThrIleArgLeu---ThrAsnLysValGluPheProPheLeu 19
Db 474 ATGGAAAGGCAATTACTTCCCAATGTTGAAAGAAATGCCCTGATTCGTTGG 533
Qy 20 ValIleuLysSerGlyGlyHisCysLeuLeuAlaLeuValGlnIleValSerAspPhe 39
Db 534 GCATTATGATTTCAGTGGACACACCAACTGTGTAAGTTGACCGCTTGGCAATAC 593
Qy 40 LeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArg 59
Db 594 GAACCTACTGGGAAATCAATTATGATGCTGCCGGTGAACCTTTGACAAACAGGCAGA 653
Qy 60 ArgLeuSerLeuIleHisAspProGlyCysSerThrMetSerGlyLysValAlaIleGlu 79
Db 654 TCACTCGGTTG---GATTACCT-----GCCGTGTAGGATGTC 692
Qy 80 HisLeuAlaLysGlnIleGlyAsnArgPheHisAspAspIleLysProProLeuHisAla 99
Db 693 AATTATGCCGAATCCGACGCCCAATCGTTTAAATTCCTGTCATGACCGACAGA 752
Qy 100 LysAsnCysAspPheSerThrGlyLeuGlnHisValThrAspLysIleIleMetLys 119
Db 753 CCGGACTGGATTTTCAGTTTCCGTTTAAACCTTTGCTGCGAATACGATTAAAGCC 812
Qy 120 LysGluLysGluGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAla 139
Db 813 AATCTTAATGAATAATGTGAATCGATGAGCA-----ACCAATGCCATATTGCC 863
Qy 140 AlathrValGlnHisThrMetAlaCysHisIleuValLysArgThrHisArgAlaIleLeu 159
Db 864 CAGCATTCCAACAGCCGTC-----GTTGATCAATTATTAAT 902
Qy 160 PheCysLysGlnArgAspLeuProGlnAsnAsnAlaValIleuValAlaSerGlyIle 179
Db 903 AATGCAAG---CGAGCGTTAGAGCAAAACCGGCTTAATACGTTAATGCGACGGCGC 959
Qy 180 ValAlaSerAspPheTyrIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCys 199
Db 960 GTAAAGGCCAATAAACAAATTACGACGACCTTCCGAATATGTAAGAAAATTAAGGCC 1019
Qy 200 ThrLeuLeuCysProProArgLeuCysThrAspAsnGlyIleMetIleAlaTrpAsn 219
Db 1020 GAAGTATTTTACCTTCGCCCACAATTTTGCATCTGACACGCGCAATGATTCGTTACT 1079
Qy 220 GlyIleGluArgLeuArg 225
Db 1080 GCGTTTCTTCGCTTAAA 1097

```

Search completed: November 11, 2005, 03:03:04
Job time : 1105.51 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 10, 2005, 16:53:13 ; Search time 3784.43 Seconds
(without alignment)
2685.516 Million cell updates/sec

Title: US-10-649-273-2_COPY_148_414
Perfect score: 1385
Sequence: 1 MEAHALITRLTNKVFPPPLV.....DISKVGASIKVQLKMEI 267

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n_model -DEV=xlp
-O=/cgm2.1/USPFO_spool_p/US10649273/runat_02112005_091338_15562/app_query.fasta_1.1429
-DB=EST -QPM=fastcap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-DOCFALGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pct -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10649273 @CGN 1 1 8076 @runat 02112005_091338_15562 -NCPU=6 -ICPU=3
-NO_MMAP -IARSEQOUTR -NEG_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOC
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST: *
1: gb_est1: *
2: gb_est2: *
3: gb_est3: *
4: gb_est4: *
5: gb_est5: *
6: gb_est6: *
7: gb_est7: *
8: gb_est8: *
9: gb_est9: *
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1263	91.2	870	5	BQ423651 AGENCOURT
2	1224	88.4	2284	3	AK045669 Mus muscu
3	1216	87.8	1622	3	AK011265 Mus muscu
4	1097	79.2	640	5	B0636028 hnd03d11.y
5	1075	77.6	852	3	BX391919 BX391919
6	1071	77.3	1503	3	BCU30671 Mus muscu
7	1053	76.0	701	2	BR740611 B0961028
8	944	68.2	922	5	B0961028 AGENCOURT
9	925	66.8	658	7	CF362328 829596 MA

10	916	66.1	822	7	CF257246	CF257246 pha008_90
11	838	60.5	637	7	CK941819	CK941819 4065407 B
12	830	59.9	597	6	CB272391	CB272391 ma57912
13	819.5	59.2	792	5	BU403563	BU403563 604138456
14	816	58.9	730	7	CN823245	CN823245 Oa.sp1bn
15	789	57.0	545	1	AV602901	AV602901 AV602901
16	786	56.8	866	5	BU127463	BU127463 603114407
17	784	56.6	490	6	CB852881	CB852881 UI-CF-FNO
18	749	54.1	736	7	CK361915	CK361915 AGENCOURT
19	749	54.1	1173	6	CD508917	CD508917 CDA93 -E05
20	746	53.9	723	5	BU261251	BU261251 603502215
21	722	52.1	812	5	BU264489	BU264489 603788202
22	717	51.8	634	2	AM601179	AM601179 RCI-BT025
23	708	51.2	484	1	AJ670918	AJ670918 AJ670918
24	708	51.1	736	6	CA057753	CA057753 sea1x5b54
25	707.5	51.1	701	5	BU621780	BU621780 UI-H-FL1
26	704	50.8	749	7	CK982692	CK982692 4115311 B
27	698	50.4	696	5	BU302606	BU302606 603739448
28	692	50.0	579	7	CK819035	CK819035 1106d05.Y
29	691.5	49.9	789	5	BU242187	BU242187 603781023
30	683	49.3	909	5	BX756548	BX756548 BX756548
31	682	49.2	548	7	CO880741	CO880741 BoyGen.09
32	671	48.4	682	2	BB043703	BB043703 BB043703
33	668	48.2	706	5	BU202465	BU202465 603940502
34	666	48.1	869	5	BX754527	BX754527 BX754527
35	647	46.7	919	7	CF407294	CF407294 CH3#047 B
36	646	46.6	878	5	BX776940	BX776940 BX776940
37	645	46.6	533	4	BM126453	BM126453 iE06d05.Y
38	638	46.1	1082	2	BX359023	BX359023 BX359023
39	625	45.1	506	2	BF415802	BF415802 UI-R-CA1-
40	618.5	44.7	879	5	BU256052	BU256052 603746167
41	614	44.3	861	5	BU246158	BU246158 603779906
42	611	44.1	1171	5	BU261605	BU261605 603501763
43	604	43.6	424	1	AA273921	AA273921 vP99C03.r
44	602	43.5	413	1	AA589724	AA589724 v178C10.r
45	600	43.3	863	5	BU376295	BU376295 603808890

ALIGNMENTS

RESULT 1
BQ423651 870 bp mRNA linear EST 23-MAY-2002
LOCUS BQ423651
DEFINITION AGENCOURT 7790948 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:6065828
5', mRNA sequence.
ACCESSION BQ423651 GI:21118966
VERSION BQ423651.1
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapds@mail.nih.gov
Tissue Procurement: ATCC/DCTD/DRP
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LLM13342 row: 1 column: 21
High quality sequence stop: 710.
location/Qualifiers
1..870
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6065828"

FEATURES

source

/tissue type="melanotic melanoma"
/lab host="DH10B (phage-resistant)"
/clone lib="NIH MGC 72"
/note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NciI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2 kb. Library constructed by Life
Technologies."

ORIGIN

Alignment Scores:

Pred. No.:	1.9e-136	Length:	870
Score:	1263.00	Matches:	248
Best Similarity:	98.42%	Conservative:	1
Best Local Similarity:	98.02%	Mismatches:	2
Query Match:	91.19%	Indels:	2
DB:	5	Gaps:	0

US-10-649-273-2_COPY_148_414 (1-267) x BQ423651 (1-870)

Qy 1 MetGluAlaHisAlaLeuThrIleArgLeuPheAsnLysValGluPheProPheLeuVal 20
Db 112 ATGAGGCTCATGCACTTACTATTAGTTGACCAATTAAGTAATTCCTTTTATGTT 171
Qy 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
Db 172 CTTTGTATTTCTGAGATGCACTGCTCTGTGGCATTAGTTCAAGAGATTTTCG 231
Qy 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60
Db 232 CTTCTTGAAAGCTTTTGACATAGCACAGGTGACATGCTTGACAGGTGCAAGAA 291
Qy 61 LeuSerLeuIleLysHisProGlnCysSerPheMetSerGlyGlyValAlaIleGlnHis 80
Db 292 CTTTCTTAAATAAACATCCAGAGTCTCCACCATGAGTGGTGGAAAGCCATTAGAACAT 351
Qy 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 100
Db 352 TTGGCCAAACAGAGAAATGATTTTCATTTTGACATCAAAACCTCCCTTGATCATGCTA 411
Qy 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120
Db 412 AATTGTGATTTTCTTTTACTGAGCTTCAACAGCTTACTGATTAATAATATGAAAAAG 471
Qy 121 GluLysGluGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaAla 140
Db 472 GAAAAAGAGAAAGGATTTGAGAAAGGGCAAAATCCGTCTTCAGCAGCAACATTTGCTGC 531
Qy 141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160
Db 532 ACAGTACAGCAACAATGGCATGTCATCTTGTAAGAAAGAACACATCGGGCTATTTCTGTT 591
Qy 161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyVal 180
Db 592 TGTAAAGAGAGAGACTTGTACTCAAAATATATGACATCTGCTGTCATCTGCTGCTGTC 651
Qy 181 AlaSerAsnPheIleThrArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThr 200
Db 652 GCAAGTAACTTCTAATATCCGAGACTCTCGAAATTTTAAACAACCGCAACAGATGCACT 711
Qy 201 LeuLeuCysProProArgLeuLeuCysThrAspAsnGlyIleMetIleAlaTrpAsnGly 220
Db 712 TTGGTGTCTCTCCCTCCAGACTATGACATGATTAATGGCATTTATGATTCAGAAATGCT 771
Qy 221 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTrpGlu 240
Db 772 ATTGAAGAACTAATCGTCTGCTTGGGCACTTTTATCATGATGATGAGGATCCGCTATGAA 831
Qy 241 -ProLysCysPro-LeuGlyValaAspIleSerLys 251
Db 832 ACCAAATGCTCCTCTGAGATGACATATCAAAA 866

RESULT 2
AK045669

LOCUS AK045669 2284 bp mRNA linear HTC 03-APR-2004
DEFINITION Mus musculus adult male corpora quadrigenina cDNA, RIKEN
full-length enriched library, clone:B230219017 product:similar to
PUTATIVE SIALOGLYCOPROTEASE TYPE 2 [Homo sapiens], full insert
sequence.
ACCESSION AK045669
VERSION AK045669.1 GI:26337528
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE 1
AUTHORS Carninci, P. and Hayashizaki, Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636

REFERENCE 2
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159

REFERENCE 3
AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Kono, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
TITLE RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multichannel sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861

REFERENCE 4
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the
PANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE 5
AUTHORS The PANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)

REFERENCE 6
AUTHORS Adachi, J., Aizawa, K., Akimura, T., Arawaka, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.
TITLE Direct Submersion
JOURNAL Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.jp,
url: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)

COMMENT CDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken

Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL: <http://genome.gsc.riken.jp/>
URL: <http://fantom.gsc.riken.jp/>.
Location/Qualifiers

FEATURES

SOURCE

1..2284
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:B30219017"
/db_xref="taxon:10090"
/clone="B230219017"
/sex="male"
/tissue_type="corpora quadrigemina"
/clone_id="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
221..1465
/note="unamed protein product; putative
similar to PUTATIVE SIALOGLYCOPROTEASE TYPE 2 [Homo
sapiens] (SPTP|O9H4B0, evidence: FASTY, 80%ID, 100%length,
match=1242) "
/codon_start=1
/protein_id="BAC32450.1"
/db_xref="GI:26337529"
/translation="MLMLRTGAIPKPKSKVYGLRRFVHPRTLSCKLVLGIET
SCDPTGAIVDEITVGEALHSGTOVHLKGTGIVPVAQOLRENIQIVETISAC
RTSPDSALITIKPGLALSLGVSLSLOLVNPKPFRPHHMEALILRTLNK
VEPPFVILLISGHCLELAVGVSPFLIGSLDIPAGMDLKVARRLSLIPKPEST
MSGKALIAAVQAHFHTINPQNAKNCDFSTGLQHTIDKLIHKEKEGLEKQ
QILSSADIAAVQAHFHTINPQNAKNCDFSTGLQHTIDKLIHKEKEGLEKQ
KALEIVANATOCCLCPPLRLCTDNGIMIAMNGIRIRLRLAGLVDVEDIRYEPKPL
GVDSIRVEABAIVKPRKML"

CDS

polyA_signal
2262..2267
/note="putative"
polyA_site
2284
/note="putative"

ORIGIN

Alignment Scores:
Pred. No.: 2.75e-131 Length: 2284
Score: 1224.00 Matches: 23
Percent Similarity: 92.51% Conservative: 14
Best Local Similarity: 87.27% Mismatches: 20
Query Match: 88.38% Indels: 0
Gaps: 0

US-10-649-273-2_COPY_148_414 (1-267) x AK045669 (1-2284)

QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThraenLysValGluPheProPheLeuVal 20
Db 662 ATGGAGGCTCAGCGACTGACTATTAGGCTCACCAATTAAGATATTCCTTTTAACTT 721
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGInGlyValSerAspPheLeu 40
Db 722 CTTTGTGATTTCGGCGGTCACTGCTGTGGCATTAAGTCCAAAGCGTTCCTCCATTTCCG 781
QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60
Db 782 CTCCTGGAGAGTCTTGGACATAGCACAGCGGACATGCTTGACAAAGGTGGCAAGAAGA 841
QY 61 LeuSerLeuIleLysHisProGluLysSerThrMetSerGlyGlyLysAlaIleGluHis 80
Db 842 CTTTCTTTAATCAACATCCAGAAAGTCTCAAGAGTGGGGAAGAGCTATAGAACAG 901
QY 81 LeuAlaLysGInGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 100
Db 902 TTGGCCAAAGCGAAATAGATTCCATTTTACTATCAATCCCTTAGCAGAAATCTAAG 961
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValAlaThrAspLysIleIleMetLysLys 120
Db 962 AATTGGCAATTTTCTTTCACGGGACTTCAACATATTATCTGATTAAGCTAATAACACAAG 1021

QY 121 GluLysGluGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaSerIleAlaAla 140
Db 1022 GAAAAAGAAAGAGGCAATTCAGAGGGGCAAAATCCTTCATCAGCTGACAGATGCTGCT 1081
QY 141 ThrValGlnHisThrMetAlaCysHisIleLeuValLysArgThrHisArgAlaIleLeuPhe 160
Db 1082 GCGGTACAGCATGCAACAGCGTCCACTGGCAAAAAGAACACATCCGCGTATTCGTCTT 1141
QY 161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyVal 180
Db 1142 TGCAGACGAAAAATTTGCTCTCTCTCCAGCTACCGCATTAATTAATTCATTCGAGGTCTT 1201
QY 181 AlaSerAsnPheTyrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThr 200
Db 1202 GCAAGTAACTTGTACATCCGAAAGACATTGGAATGTCGCAATGCAAGCAGGTGACG 1261
QY 201 LeuLeuCyProProProArgLeuCyThrAspAsnGlyIleMetIleAlaTPAsnGly 220
Db 1262 TTGTTGTCTCCACCTCCAGACATGTCACACTGACACAAATGCAATGATTCATGCAATGGA 1321
QY 221 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrGlu 240
Db 1322 ATTGAAGATTACGCTGCTGCTGGCGGTTTACATGATGTAGAAACATCCATGGA 1381
QY 241 ProLysCyProLeuGlyValAspIleSerLysGlyValGlyGlnAlaSerIleLysVal 260
Db 1382 CCAAAATGCTCTTTCGAGTACATATCCAGAAAGTTGCACAAAGCTGCATTAAGTA 1441
QY 261 ProGlnLeuLysMetGluIle 267
Db 1442 CCGCATTAATAAATGCACCTT 1462

RESULT 3
AK011265
LOCUS 1622 bp mRNA linear HTC 03-APR-2004
DEFINITION Mus musculus 10 days embryo whole body cDNA, RIKEN full-length
enriched library, clone:261001M19 product:similar to PUTATIVE
SIALOGLYCOPROTEASE TYPE 2 [Homo sapiens], full insert sequence.
ACCESSION AK011265
VERSION AK011265.1 GI:12847275
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Komano, H., Akiyama, J., Nishi, K., Katsunari, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matsubara, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--194-Format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)

TITLE
JOURNAL
MEDLINE
PUBMED
AUTHORS
REFERENCES
1 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Komano, H., Akiyama, J., Nishi, K., Katsunari, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matsubara, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--194-Format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)

TITLE
JOURNAL
MEDLINE
PUBMED
AUTHORS
REFERENCES
1 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Komano, H., Akiyama, J., Nishi, K., Katsunari, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matsubara, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--194-Format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)


```
/cell_type="HELA CELLS COT 25-NORMALIZED"  
/cell_line="HELA"  
/clone_lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo(dT)  
primer. Five prime end enriched, double-strand cDNA was  
digested with Not I and cloned into the Not I and EcoR V  
sites of the pCMVSPORT 6 vector. Library was normalized."
```

ORIGIN

Alignment Scores:

Pred. No.:	1.65e-114	Length:	852
Score:	1075.00	Matches:	208
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	77.62%	Indels:	0
DB:	5	Gaps:	0

US-10-649-273-2_COPY_148_414 (1-267) x BK391919 (1-852)

```
Qy 1 MetGluAaHsAlaLeuThrTlLeArgLeuThraenLysValGluPheProPheLeuVal 20  
Db 626 ATGAGAGCTCATGCACTTACTTATAGTTGACCAATTAAGTAAATTTCTTTTATGTT 567  
Qy 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40  
Db 566 CTTTGTGATTTCTGGAGGTCACGTCCTGTTGGCATTAGTTCAAGAGATTTCAGATTTCTG 507  
Qy 41 LeuLeuGlyLysSerLeuAspTlLeAlaProGlyAspMetLeuAspLysValAlaArgArg 60  
Db 506 CTTTGTGAAAGCTCTTGGACATAGCACAGGTGACATGCTTGACAAAGTGGCAAGAGA 447  
Qy 61 LeuSerLeuIleLysHisProGlyCysSerThrMetSerGlyGlyValAlaIleGluHis 80  
Db 446 CTTTCTTAAATTAACAATCCAGAGTCTCCACCATGATGATGGTGGGAAAGCCATAGAACT 387  
Qy 81 LeuAlaLysGlnGlyAsnArgPheHisAspIleLysProProLeuHisHisAlaLys 100  
Db 386 TTGGCCAAACAAAGAAATAGATTTCAATTTTGAATCAAAACCTCCCTGCATCATGCTAA 327  
Qy 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120  
Db 326 AATTGTGATTTTCTTTTACTGAGCTTCAACCGTTACTGATTAATAATATGAAAAAG 267  
Qy 121 GluLysGlnGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAla 140  
Db 266 GAAAGAGAGAAAGTATTGAGAGAGGCAAAATCTGTCTTCCAGACAGACATTTGCTGCC 207  
Qy 141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160  
Db 206 ACAAGTACGACACACATGATGATCTTGTGAAAAGAACACATCGGCTATTTCTGTTT 147  
Qy 161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyVal 180  
Db 146 TGTAAAGCAGAGAGACTTGTACTTCAAAATATATGACATGATGTCATCTGTGTGTGC 87  
Qy 181 AlaSerAsnPhePheTlLeArgArgAlaLeuGlnIleLeuThraenLysValGluPheProPheLeuVal 200  
Db 86 GCAAGTAACTTCTATATCCGAGAGCTCTGGAATTTTAAACAAACGCAACACAGTGACT 27  
Qy 201 LeuLeuCysProProArgLeu 208  
Db 26 TTGTTGTCTCTCTCTCCAGACTA 3
```

```
RESULT 6  
LOCUS BC030671 1609 bp mRNA linear HTC 19-NOV-2003  
DEFINITION Mus musculus O-sialoglycoprotein endopeptidase-like 1, mRNA (cDNA  
clone IMAGE:1226118), containing frame-shift errors.  
ACCESSION BC030671  
VERSION BC030671.1 GI:21040459  
KEYWORDS HTC.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus
```

REFERENCE
AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 1609)

Klausner, R.D., Collins, F.S., Wagner, L.H., Derge, J.G., Strausberg, R.L., Feingold, R.A., Grouse, L.H., Derge, J.G., Altschul, S.F., Zeeberg, B., Burow, K.H., Scheffer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, U., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullan, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Huiyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shcherchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butcherfield, Y.S., Krzywinski, M.I., Skalska, U., Smallus, D.E., Scherch, A., Schein, J.R., Jones, S.J. and Marra, M.A.

Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

22388257
12477932
2 (bases 1 to 1609)

Strausberg, R.
Direct Submission
Submitted (20-May-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK
COMMENT
Contact: MGC help desk
Email: cgabbs-remail.nih.gov
Tissue Procurement: Marcello Bento Soares, Ph.D.
cDNA Library Preparation: M. Bento Soares, University of Iowa
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)
DNA Sequencing by: Institute for Systems Biology
http://www.systemsbio.org
contact: amadan@systemsbiology.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/BLN at: <http://image.llnl.gov>
Series: IRK Plate: 66 Row: e Column: 10
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 21312463
This clone has the following problem: frame shifted.

FEATURES
source

1..1609
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:1226118"
/issue_type="Thymus gland, mouse"
/clone_lib="Soares_thymus_2N2MT"
/lab_host="DH10B"
/note="Vector: pT73-Pac"

ORIGIN

Alignment Scores:

Pred. No.:	1.21e-113	Length:	1609
Score:	1071.00	Matches:	211
Percent Similarity:	83.52%	Conservative:	12
Best Local Similarity:	79.03%	Mismatches:	20
Query Match:	77.33%	Indels:	24
DB:	3	Gaps:	1

US-10-649-273-2_COPY_148_414 (1-267) x BC030671 (1-1609)

QY 1 MetGUAlAHISAlAleuthr1leargleuthrAsnlyValGluPheProPheleuVal 20
Db 665 ATGAGAGCTCAAGCACTACTATTAGGCTCACCAATAAGTAGAATTTCTTTTATAGTT 724
QY 21 LeuLeu1leSerg1yG1yHIScYsleuLeuAlaleuValGlnGlyValSerAppheleu 40
Db 725 CTTTGGATTTCGGGGTCACTGCTGTTGGCATTAGTCCAAAGGTGTTCCATTTCCTG 784
QY 41 LeuLeuG1yLysSerLeuAsp1leAlaProG1yAspMetLeuAsp1yValAlaArg 60
Db 785 CTCCTTGGAGATCTTTGGACATAGCACCGACGACATCTTGACAGGTGGCAAGACA 844
QY 61 LeuSerLeu1leYsHISProG1yCysSerThrMetSerg1yG1yLysAla1leGluHIS 80
Db 845 CTTTCTTAATCAAAACATCCAGAAATGTTCTCAATGAGTGCGGAAAGACTATAGAACAG 904
QY 81 LeuAla1ySGlnG1yAsnArgPheHISpHeAsp1leYsProProLeuHISHisAlaYs 100
Db 905 TTGGCCAAAGACGGAATAGATTCCATTACTATCAATCCACTATGCAAGATGCTAG 964
QY 101 AsnCyAspPheSerPheThrG1yLeuGlnHISValThraSp1y1le1leMetLys 120
Db 965 AATTGCATTTTCTTTCAAGGACTTCAACATATTACTGATTAAGCTAATACACACAG 1024
QY 121 GluYsGlnG1yG1y1leG1yLysG1yGln1leLeuSerSera1AlaAsp1leAla 140
Db 1025 GAAAGAGAGAGGATTCAGAGAGGGCAATCTGTATCACTGACGACATTCGTCT 1084
QY 141 ThrValGlnHISThrMetAlaCySHISleuVal1ySargThrHISArgAla1leLeuPhe 160
Db 1085 GCGGTACAGCATGCAACAGCGGCACTTCGCAAGAACACATCGCGCTATTCGTCT 1144
QY 161 CysLysGlnArgPheLeuProGlnHISAsnAlaVal1leValAlaSer1yG1yVal 180
Db 1145 TGCAGACGAGAAATTTCTCTCTCCAGCTAACGAGATTAAGTTGATCTGAGGCTGT 1204
QY 181 AlaSerAsnPheYr1leArgArgAlaLeuGln1leLeuThraSnAlaThrGlnCysThr 200
Db 1205 GCAAGTAACTTGATCATCCGAAAGCATTTGGAAATTCGCAATGCAAGCGACAG 1264
QY 201 LeuLeuCySProProProArgLeuCySThrAspAsnG1y1leMet1leAlaTrpAsnG1y 220
Db 1265 TTGTTGTCCTCACCCTCAAGACTGTGCACTGCAATGGCATCATGATTGCA 1315
QY 221 1leG1yArgLeuArgAlaG1yLeuG1y1leLeuHISAsp1leGlnG1y1leArgTyGln 240
Db 1315 1315
QY 241 ProLysCySProLeuG1yValAsp1leSerLysGlnVal1yG1yAlaSer1leLysVal 260
Db 1316 ---TGATGTCCTCTTGGAGTACATATCCAGAGAGTTGCCAAGCTGCCATTAAGTA 1372
QY 261 ProGlnLeuLysMetGln1le 267
Db 1373 CCGCGATTAAAAATGCACTT 1393
RESULT 7
BE740611
LOCUS 601595739P1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3949640 5',
DEFINITION mRNA sequence.
ACCESSION BE740611.1 GI:10154603
VERSION BE740611.1
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 701)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: DCTD/DRP
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at: image.lnl.gov
plate: LNC814 row: n column: 09
High quality sequence stop: 701.
Location/Qualifiers
1..701
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3949640"
/issue_type="adenoecarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: ovary; Vector: pOTB7; Site:1: XhoI; Site_2:
EcoRI; CDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCAAGAG(C). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
ORIGIN
Alignment Scores:
Pred. No.: 4,6e-112 Length: 701
Score: 1053.00 Matches: 206
Percent Similarity: 99.054 Conservative: 3
Best Local Similarity: 97.634
Query Match: 76.038 Indels: 2
DB: 2 Gaps: 0
US-10-649-273-2_COPY_148_414 (1-267) x BE740611 (1-701)
QY 1 MetGUAlAHISAlAleuthr1leargleuthrAsnlyValGluPheProPheleuVal 20
Db 71 ATGAGAGCTCAAGCACTACTATTAGGCTCACCAATAAGTAGAATTTCTTTTATAGTT 130
QY 21 LeuLeu1leSerg1yG1yHIScYsleuLeuAlaleuValGlnGlyValSerAppheleu 40
Db 131 CTTTGGATTTCGGGGTCACTGCTGTTGGCATTAGTCCAAAGGTGTTCCATTTCCTG 190
QY 41 LeuLeuG1yLysSerLeuAsp1leAlaProG1yAspMetLeuAsp1yValAlaArg 60
Db 191 CTTTGGAGATCTTTGGACATAGCACCGACGACATCTTGACAGGTGGCAAGACA 250
QY 61 LeuSerLeu1leYsHISProG1yCysSerThrMetSerg1yG1yLysAla1leGluHIS 80
Db 251 CTTTCTTAATCAAAACATCCAGAAATGTTCTCAATGAGTGCGGAAAGACTATAGAACAT 310
QY 81 LeuAla1ySGlnG1yAsnArgPheHISpHeAsp1leYsProProLeuHISHisAlaYs 100
Db 311 TTGGCCAAAGACGGAATAGATTCCATTACTATCAATCCACTATGCAAGATGCTAG 370
QY 101 AsnCyAspPheSerPheThrG1yLeuGlnHISValThraSp1y1le1leMetLys 120
Db 371 AATTGCATTTTCTTTCAAGGACTTCAACATATTACTGATTAAGCTAATACACAG 430
QY 120 yGlnG1yGlnG1y1leG1yLysG1yGln1leLeuSerSera1AlaAsp1leAla 140
Db 431 GGAAGAGAGAGGATTCAGAGAGGGCAATCTGTCTTCAAGACGACATTCGTCTG 490
QY 140 1aThrValGlnHISThrMetAlaCySHISleuVal1ySargThrHISArgAla1leLeuP 160
Db 491 CCACAGTACAGCAACCAATGCAATTCATCTTGTAAAGAAACACATCGGCTATTCCT 550
QY 160 heCyLysGlnArgPheLeuProGlnHISAsnAlaVal1leValAlaSer1yG1yVal 180

Db 551 TTGTGAAGCAGAGACTGTTTACTCTCAAAATATATCGACTGTTGCATCTGTCGTG 610
Qy 180 aAlaSerAspPheTyrIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysT 200
Db 611 TCGCAAGTAACCTTCTATATCCGACAGACTCTGGAAATTTTAAACCAACGACAGTGA 670
Qy 200 hrLeuLeuCysProProProArgLeuCys 209
Db 671 CTTGTGTGTGTCCTCTCCAGACTATGC 699
RESULT 8
LOCUS BQ961028 922 bp mRNA linear BST 21-AUG-2002
DEFINITION AGENCOURT 8863711 NIH_MGC_18 Homo sapiens cDNA clone IMAGE:6423902
5', mRNA sequence.
ACCESSION BQ961028
VERSION BQ961028.1 GI:22376506
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 922)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCTD/DTF/Cazdar
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LNCM2603 row: d column: 15
High quality sequence stop: 584.
Location/Qualifiers
1..922
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6423902"
/tissue_type="large cell carcinoma"
/lab_host="RD10B (phage-resistant)"
/clone_lib="NIH_MGC_18"
/note="Organ: Lung; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."

ORIGIN
Alignment Scores:
Pred. No.: 3.65e-99 Length: 922
Score: 944.00 Matches: 207
Percent Similarity: 89.45% Conservative: 5
Best Local Similarity: 87.34% Mismatches: 16
Query Match: 68.16% Indels: 10
DB: 5 Gaps: 3
US-10-649-273-2_COPY_148_414 (1-267) x BQ961028 (1-922)

Qy 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20
Db 207 ATGGAGGCTCARGCACTTACTATTAGGTGGACCAATAAGATTTCTTTTATAGTT 266
Qy 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
Db 267 CTTTGAATTTCTGGAGGTCACGTGCTGTGGCATTAGTTCAAGAGATTTCACATTTTCTG 326

Qy 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60
Db 327 CTTCTGGAAAGACTCTTGGACATGACACAGGAGACATCTTGAACAAGGCGAAGAAG 386
Qy 61 LeuSerLeuIleLysHisProGluCys-SerThrMetSerGlyGlyValAlaIleGluHis 80
Db 387 CTTTCTTAAATATAACATCCAGAGTGCNTCCACCAAGTGTGGGAGAGCCATTAACCA 446
Qy 80 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 100
Db 447 TTGGCCCAACACAGAAATAGATTTCATTGATTCATCAACCTCCCTTGATATAGCTTA 506
Qy 100 sAsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120
Db 507 AAATTGTATTTTCTTTTACTGACCTTCAACAGCTTACTGATTAATATATATGAAAAA 566
Qy 120 sGluLysGluGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaI 140
Db 567 GGAAAAAGAGGAAGGATTTAGAGAAAGGCAAAATCCGTCTTCAGCAGCAGACATTCCTGC 626
Qy 140 aThrValGlnHisSTHMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuP 160
Db 627 CACAGTACGACACACATGCAATGCAATGTCATCTTGTAAGAAAGAAA-CATCGGAGCTATTTCTGTT 685
Qy 160 eCysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSer-GlyGlyLys 180
Db 686 TTGTGAAGCAGAGACTTGTAACCTGTAACCTCAAAATATATGACGATGCGATCTTGCGGGGG 745
Qy 180 aAlaSerAspPheTyrIleArgArgAlaLeuGluIleLeuThr-AsnAlaThrGlnCys 199
Db 746 TCGCAAGTAACCTTATATATCCCGCAAACTCTGGAAAAATTTACAAACCCACACACGCGGG 805
Qy 200 Thr-LeuLeuCysProProProArgLeuCysThr-AspAsnGlyIleMetIleAlaT 218
Db 806 ACCTTGTTGGGTCCCTCCCAACTATGCACTGATATATGCG---CATTAATTGATGCG 862
Qy 219 -----AsnGlyIleGluArg---LeuArgAlaGlyLeuGly 229
Db 863 TCGGGAAGGAAATTGAAAAAATAATACNTGCTCGCTTGGGG 903
RESULT 9
CF362328/c 658 bp mRNA linear EST 25-AUG-2003
LOCUS CF362328/8 829596 MARC 3pig Sus scrofa cDNA 3', mRNA sequence.
DEFINITION CF362328
ACCESSION CF362328
VERSION CF362328.1 GI:34161882
KEYWORDS EST.
SOURCE Sus scrofa (pig)
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE 1 (bases 1 to 658)
AUTHORS Smith,T.P.L., Freking,B.A., Ford,J.J., Vallet,J.L., Fox,J.,
Wise,T.A., Nonneman,D.J., Wray,J.E. and Keeler,J.W.
A second set of porcine ESTs from a pooled-tissue normalized
library
JOURNAL Unpublished (2003)
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing: Bases called with phred v0.020425.c and
trimmed with the aid of the trim_al option. Vector identified with
cross match v0.990329.
plate: SRG8015 row: H column: 12
Seq primer: TAGAAGCACAATCGACG.
FEATURES
source
1..658
/organism="Sus scrofa"
/mol_type="mRNA"

/db_xref="taxon:9823"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_id="MARC 3P1G"
/note="Vector: pCDNA3.1; Site 1: EcoRI; Site 2: NotI;
Library made with RNA pooled from multiple tissues
including brain, liver, muscle, placenta/endometrium,
ovary, testes, and bone marrow."

ORIGIN

Alignment Scores:

Pred. No.:	3 71e-97	Length:	658
Score:	925.00	Matches:	174
Percent Similarity:	93.97%	Conservative:	13
Best Local Similarity:	87.44%	Mismatches:	12
Query Match:	66.79%	Indels:	0
DB:	7	Gaps:	0

US-10-649-273-2_COPY_148_414 (1-267) x CF362328 (1-658)

QY 20 ValLeuLeuLeuSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPhe 39
DB 657 GTTCTTTGATATCTGGCGGTCATGCTTTGGCATTAGTAGAGAGGTTTCAGATTTT 598
QY 40 LeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArg 59
DB 597 CTGCTTCTTGACAGCTCTTGACATAGCACAGGTGACATGCTTGACAGGTGACAGCA 538
QY 60 ArgLeuSerLeuLeuLysHisProGlyLysSerThrMetSerGlyGlyValAlaIleGlu 79
DB 537 AGACTTTCTTAAATAAACATCCAGAGTGTCTCCACCATGATGTGGGAAAGCCATAAA 478
QY 80 HisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisAla 99
DB 477 CATTTGTCACAAAGGAAATTAAGTGCATTTCATTCACAACTCCCATGCAAGCTAT 418
QY 100 LysAsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLys 119
DB 417 AAAAATTTGATTTTCTTTTCTTGACCTTCAACATGTTATGATTAAGCAATATAGCAG 358
QY 120 LysGlnLysGlnGlyIleGlyLysGlyGlnIleLeuSerSerAlaAlaAspIleAla 139
DB 357 AAGGAAAGAGAGAGAGATTTGAGAAAGGGCAATCTGCTTCAGCTGCAGACATTGCT 298
QY 140 AlaThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeu 159
DB 297 GCGCAGATACAGACACAGTAGCTGCTGCATATTCACAAAGACCTCATGCTATTTCTG 238
QY 160 PheCysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGly 179
DB 237 TTTTGCACACAGAGAGCTTATTAATGTCACAAAGTAATGCAATTTGTTGATCTGAGGT 178
QY 180 ValAlaSerAsnPheTyrlleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCys 199
DB 177 GTTGCAAGTAATCTTATATATACGAAAGCTTTAGAAAGTTGACAAATGCAACAAATGC 118
QY 200 ThrLeuLeuCysProProArgLeuCysThrAspAsnGlyIleMetIleAlaTrp 218
DB 117 ACTCTGTGTGCTCTCTCCACAGCTATGACATGATATAGCATTAATGATTGATGG 61

RESULT 10

CF257246

LOCUS CF257246 822 bp mRNA linear EST 07-AUG-2003

DEFINITION pha008_g02 PHA-activated splenocytes Gallus gallus cDNA, mRNA

ACCESSION CF257246

VERSION CF257246.1 GI:33490501

KEYWORDS EST

SOURCE Gallus gallus (chicken)

ORGANISM Gallus gallus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.

REFERENCE 1 (bases 1 to 822)
AUTHORS Wittzell,H., Bed'Hom,B., Morin,V., Young,J.R., Whitaker,C.J.,
Chausse,A.M. and Zoarob,R.
TITLE A collection of chicken ESTs from activated immune cells
JOURNAL Unpublished (2003)
COMMENT Contact: Zoarob R
UPR 1983
CNRS

7 rue Guy Moquet, BP 8, 94801 Villejuif cedex, France
Tel: 33 1 49 58 35 00
Fax: 33 1 49 58 33 81
Email: zoarob@jf.cnrs.fr.
Location/Qualifiers

FEATURES

source

1..822
/organism="Gallus gallus"
/mol_type="mRNA"
/db_xref="taxon:9031"
/cell_type="Splenocytes"
/clone_id="PHA-activated splenocytes"
/note="Vector: pTriplex2"

ORIGIN

Alignment Scores:

Pred. No.:	5.75e-96	Length:	822
Score:	916.00	Matches:	175
Percent Similarity:	81.37%	Conservative:	39
Best Local Similarity:	66.54%	Mismatches:	49
Query Match:	66.14%	Indels:	1
DB:	7	Gaps:	0

US-10-649-273-2_COPY_148_414 (1-267) x CF257246 (1-822)

QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGlnPheProPheLeuVal 20
DB 35 ATGGAGGCTCAGCCACTTTCATCCATCCAGCTGACAGACAGCAATTTCTTCTTATGTT 94
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
DB 95 CTTTACTCTCGGAGGTACATGCATCTTGGCAGTAGACACAGAGACTTTCAGATTTCCTT 154
QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArg 60
DB 155 CTGCTTGACAGATCATGATATACACACACAGGTGACATGTTGATTAAGTAGCAAGAG 214
QY 61 LeuSerLeuIleLysHisProGlyLysSerThrMetSerGlyGlyLysAlaIleGluHis 80
DB 215 CTCTCTTAAATGAAACACCCGAGGTGCCACACATGCGCGGGGAAAGCAATAGAGCAC 274
QY 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisAlaLys 100
DB 275 CTGCTCAACACCGAGAGCTGCGACAGACATTCATTCAGACTTCCCATGCAACAGTATCGT 334
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120
DB 335 AACTGTGATTTTCTTTCTCGGACCTTGACAGCCTTGCAACAAAGCCATTCTTCAGAA 394
QY 121 GlnLysGlnGlyIleGlyLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaAla 140
DB 395 GAAAAAGAAAGGATTAATCAAGAAAGGGCAATCTGCTCGCTTAAGCAATCGCTGCT 454
QY 141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160
DB 455 GCTGCACAGACAGTAGTGTCTCTCATATTAATCAAGGAGACACACGAGCCATGCTCTTC 514
QY 161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyVal 180
DB 515 TGCATGAAAAACAGCATATTATACCAAAACTGCAACTGTGTTTATACAGAGAGAGTT 574
QY 181 AlaSerAsnPheTyrlleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysTrp 200
DB 575 GCAAGTATCATGATATATAGAAAGAGCTGCAGACTGCGCAAAATGCAACGTTTTCCT 634
QY 201 LeuLeuCysProProArgLeuCysThrAspAsnGlyIleMetIleAlaTrpAsnGly 220

Db	Accession	Source	Organism	Reference Authors	Title	Journal	Comment
Db	635	TTTCTGTCCTCTCCCTCCAGAGCTGTGCACCCGTAATGATGATTGCATGCAGATGCG					
Qy	221	11egluarqleuagialaglyleugilyleleuhtsaapllegluglyleargtyrslyu					
Db	695	ATTGAAGGCTTCCGCGCAGAGATGATGATTTTATATACAGTACTGATGCGATCGGCTACGAA					
Qy	241	ProlyscysProleuglyValaspl1eserlysgtvalglylualaserlleyVal					
Db	755	CCAAAGCTCCCTTGGAATTGATATTTCCAAAGATTGAA-GAGGATTCATCAAACTG					
Qy	261	Proglneu 263					
Db	814	CAAGACTA 822					
RESULT 11							
LOCUS	CK941819/c						
DEFINITION	CK941819	637 bp	mRNA	linear	EST 15-MAR-2004		
ACCESSION	4065447	BARC 10BOV Bos	taurus	cdna	clone 10BOV12_F24_3', mRNA		
VERSION	CK941819						
KEYWORDS	CK941819.1	GI:45456199					
SOURCE	EST.						
ORGANISM	Bos taurus	(cow)					
REFERENCE	Bukaryocia, Metazoa, Chordata, Craniata, Vertebrata, Euteleostomi;						
AUTHORS	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;						
TITLE	Bovine; Bos.						
JOURNAL	1 (bases 1 to 637)						
COMMENT	Sonstegard, T.S., Van Tassel, C.P., Matukumalli, L.K., Harhay, G.P., Boesik, S., Rubenfield, M. and Gasbarre, L.C. Production of EST from cdna libraries derived from immunologically activated bovine gut unpublished (2004)						
	Contact: Tad S. Sonstegard						
	Bovine Functional Genomics Laboratory						
	Animal and Natural Resources Institute						
	Bldg. 200 Rm2A BARC-East, Beltsville, MD 20705, USA						
	Tel: 3015048416						
	Fax: 3015048414						
	Email: tads@anri.barc.usda.gov						
	Single pass sequencing. Bases called and trimmed with phred						
	0.000925 using options -trim_alt "-trim.fasta. Vector identified						
	by cross match using options -mismatch 12 -mismatch 18						
	Plate: 12 row: P column: 24						
	Seq primer: AGCGATACCAATTCACACAGG						
	High quality sequence stop: 637.						
FEATURES	Location/Qualifiers						
Source	1..637						
	/organism="Bos taurus"						
	/mol_type="mRNA"						
	/strain="Holstein"						
	/db_xref="taxon:9913"						
	/clone="10BOV12_F24"						
	/sex="Male"						
	/tissue_type="Pooled"						
	/dev_stage="Multiple"						
	/lab_host="DH10B T1 phage resistant"						
	/clone_id="BARC 10BOV"						
	/note="Organ: Small Intestine; Vector: pagen-1; Site:1; EcorV; Site:2: NotI; Equimolar amounts of mRNA extracted						
	from proximal jejunums of 18 and 21 wk old steers, and						
	distal ileums of 14 day old calves; proximal jejunum						
	exposed to C. oncophora for 3 and 6 weeks, and distal						
	ileum exposed to C. parvum for 7 days"						
ORIGIN							
Alignment Scores:							
Pred. NO.:	5.09e-87	Length:	637				
Score:	838.00	Matches:	153				
Percent Similarity:	92.47%	Conservative:	19				
Dist Local Similarity:	82.26%	Mismatches:	14				
Query Match:	60.51%	Indels:	0				

DB:	7	Gaps:	0
US-10-649-273-2_COPY_148_414 (1-267)	x	CK941819 (1-637)	
Oy	82	AlaIysGInGInIyaEnAtrPheNi:sPheAspIleIysPProPoleuNi:SHiSaIaIySaSn	101
Db	632	GCcAAACAGGGAAATAGATTGATTTGATTTCAGGCTCCCAAGCAACGGCTMAAAT	573
Oy	102	CysAspPheSerPheThrGlyLeuGlnHisValThraPlySileIleEmetIyLeuGlu	121
Db	572	TGTAATTTTCTTTTCTGACCTTCAACACCTTTTGTATGAAGAATCAATGCCAAAGCAA	513
Oy	122	LyseGInGInGlyIleGlyIlyuysGInIleIeuserSerAlaIAspIleAlaIaIaThr	141
Db	512	AAAGAGAGAGATATCGACGAGGGCAGGCTCTGCTTCTCGCTGACAGCATTTGCTTGGC	453
Oy	142	ValGlnHisThrMeCaIaCySHiSleuValIySaTrgThriSaTgaIaIleLeuPheCys	161
Db	452	GTGCAGCAGCACCGGTGGCTGCCCATTTGCCAAAAGAACACATCGTCTTCTTCTGGC	393
Oy	162	LyseGInIyrgAspleuLeuProGInAsnAsnAlaValIleuValaIasergIyIyVala	181
Db	392	AAGCAGAGAGGCTCTTAACGTCAAGATVACACGATACGTGTATCTGAGAGGCTGCCA	333
Oy	182	SerAsnPheTrIleatrgaTgaIaIleuGInIleIeunThraSaIaIaThnGInCysThrIeu	201
Db	332	AGTAACTTATATATCCGAAAGCCCTGAAATTTGTACCAATGTACACAGTGCACGTGCAC	273
Oy	202	LeuCySPProPProIaTgIeucYsThraSasnGlyIleEmetIleAlaTrrAsnGlyIle	221
Db	272	CTGTGCCCCCCCCCAGACTTTGCACACTGCACACGCGTATGATTGCATGGAATGGTGT	213
Oy	222	GluIyrgIeuaTgaIagIyLeuGlyIleIeunHisaspIleGInGlyIleatrgIyGluPro	241
Db	212	GAAAGACTACGTGCTGGCTGGCATTTTACACAAACAGAGGCATCCGCTAGGAACA	153
Oy	242	LyseCySProleuGlyValaAspIleSerIySeGInuValGlyGInaIaSerIleIyValaPro	261
Db	152	AAAGTCTCTCTTGGAGTATATATCAAAAGATGTGAGAGCTGTATATAAAGTGCAC	93
Oy	262	GInIeuIyMetGInuIle 267	
Db	92	AGATTAAAAATGAAGATT 75	
RESULT 12			
CB272391	597 bp	mRNA	linear EST 24-FEB-2003
LOCUS	ma157612.v1	McCarrey Eddy spermatoocytes Mus musculus cDNA clone	
DEFINITION	IMAGE:6445750 5' similar to TR:Q5VMD6 Q5VMD6 CG14231 PROTEIN. ''	mRNA sequence.	
ACCESSION	CB272391		
VERSION	CB272391.1	GI:28462714	
KEYWORDS	EST.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 597)		
AUTHORS	McCarrey,J., Eddy,M., Marra,M., Hillier,L., Clifton,S., Pape,D., Martin,J., Wylie,T., Dante,M., Bowers,Y., Theising,B., Gibbons,M., Ritter,E., Tsagarishevili,R., Ronko,I., Maguire,L., Kennedy,S., Bennett,J., Waterston,R. and Wilson,R.		
TITLE	NIHS Mouse		
JOURNAL	Unpublished (2002)		
COMMENT	Contact: McCarrey/Eddy NIHS Mouse		
	NIHS Mouse		
	Washington University School of Medicine		
	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA		
	Tel: 314 286 1800		
	Fax: 314 286 1810		
	Email: est@watson.wustl.edu		
	Foundation for Biomedical Research, Dept. of Genetics - ex-cision		

done by E.M. Eddy, Ph.D. (National Institutes of Health, National Institute of Environmental Health Sciences).

WGI:2069710

Seq primer: primer name ambiguous
High quality sequence stop: 419.

FEATURES

source

location/Qualifiers
1..597

/organism="Mus musculus"
/mol_type="mRNA"
/strain="CD-1"
/db_xref="taxon:10090"
/clone="IMAGE:6445750"
/sex="male"
/tissue_type="spermatoocytes, pooled from multiple mice"
/dev_stage="60 day"
/lab_host="DH10B (phage-resistant)"
/clone_lib="McCarrey Eddy spermatoocytes"
/note="Organ: testis; Vector: pBluescript SK-
(Stratagene); Site_1: XhoI; Site_2: EcoRI; cDNA oligo
dt-primed [5'-(GA)10-ACTAGTCTCAGTCTTTTCTTTT-3'] and
directionally cloned using 5' linker 5'-ATTCCGACGAG-3'
and 5'-CTCGTCCG-3'. Size selection of >400bp material
gives average insert size ranging from 1-2 kb. Library was
mass excised (from lambda-unizap-XR) and retransformed
into DH10B. Library contains 98% recombinants.
References: J. Androl. 20:635-639 and Gene 25:263-269.
Library constructed and donated by J. McCarrey, Ph.D.
(Southwest Foundation for Biomedical Research, Dept. of
Genetics); excision done by E.M. Eddy, Ph.D. (National
Institutes of Health, National Institute of Environmental
Health Sciences). Original lambda-based library is
available through ATCC, catalog #63422."

ORIGIN

Alignment Scores:

Pred. No.: 3.98e-86 Length: 597
Score: 830.00 Matches: 162
Percent Similarity: 91.80% Conservative: 6
Best Local Similarity: 88.52% Mismatches: 15
Query Match: 59.93% Indels: 0
DB: Gaps: 6

US-10-649-273-2_COPY_148_414 (1-267) x CB272391 (1-597)

QY 1 MetGlnAlaHisAlaLeuThrIleArgLeuThrAsnIleValGlnPheProPheLeuVal 20
DB 49 ATGAGGCTCAGCAGCTGACTATTAGGCTCAACAATGAAGATTCTTTTAACTT 108
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnIleValSerAspPheLeu 40
DB 109 CTTTGAATTTCTGGGGCTCAGCTGCTGTGGCATTTAGTCCAAAGGTGTTCCATTTCTG 168
QY 41 LeuLeuGlyIleSerLeuAspIleAlaProGlyAspMetLeuAspIleValAlaArgArg 60
DB 169 CTCCTTGGGAAGTCTTTGGACATAGCACAGGCGACATGCTTGGACAAGTGGCAAGAGA 228
QY 61 LeuSerLeuIleIleValHisIleProGlyCysSerThrMetSerGlyGlyValAlaIleGluHis 80
DB 229 CTTTCTTAATCAACAACATCCAGAAATGTTCTCAATGAGTGGTGGAAAGCTATACAAAG 288
QY 81 LeuAlaIleGlnGlnIleAspArgPheHisIleAspIleIleValProPheLeuHisAlaIle 100
DB 289 TTGGCCAAAGACGAAATGATGATTCATTTTACTATCAATCCATATGACAGAAATCTAG 348
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspIleIleIleMetIleVal 120
DB 349 AATTTGCATTTTCTTTTCAAGGGACTTCAACATATTACTGTAAGCTTAATACACACAG 408
QY 121 GlnIleGlnIleGlnIleGlnIleGlnIleGlnIleGlnIleGlnIleGlnIleGlnIle 140
DB 409 GAAAGAAAGAGAGCATTTGAGAGAGGGGCAATTCGTGATCAGCTGACAGACATTTGCTCT 468

QY 141 ThrValGlnHisThrMetAlaCysHisLeuValIleArgThrHisArgAlaIleLeuPhe 160
DB 469 GCGGTACAGATGCAACACAGCGTCCACTTGGCGAAAGAACACATCCGCTATTCTGTTT 528
QY 161 CysIleGlnArgAspLeuLeuProGlnIleAsnAlaValLeuValAlaSerGlyGlyVal 180
DB 529 TCGACGAGAAATTTGCTCTCTCTCCAGCTTACGAGTATTGATCTGAGAGGTGT 588

QY 181 AlaSerAsn 183
DB 589 GCAAGTAC 597

RESULT 13

BU403563

LOCUS BU403563 792 bp mRNA linear EST 27-NOV-2002
DEFINITION 604138456F1 CSEQCHN59 Gallus gallus cDNA clone CHEST967014 5', mRNA
sequence.

ACCESSION

BU403563

VERSION BU403563.1 GI:25772619
KEYWORDS EST.

SOURCE

Gallus gallus

ORGANISM

Gallus gallus

REFERENCE

AUTHORS

BOARDMAN, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,
Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
A Comprehensive Collection of Chicken CDNAS
Curr. Biol. 12 (22), 1965-1969 (2002)

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology
(UMIST)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.
Location/Qualifiers
1..792

FEATURES

source

/organism="Gallus gallus"
/mol_type="mRNA"
/strain="White Leghorn, Hixex"
/db_xref="taxon:9031"
/clone="CHEST967014"
/dev_stage="36"
/lab_host="DH10B"
/clone_lib="CSEQCHN59"
/note="Organ: limbs; Vector: pBluescript II KS(+); Site_1:
EcoRI; Site_2: NotI; This normalized library was
constructed from 1 million independent clones. cDNA
synthesis was initiated using an oligo(dT) primer, using
methylated C in the first strand synthesis reaction.
Following this first strand reaction, double-stranded cDNA
was blunt-ended, ligated to NotI adaptors, digested with
EcoRI, size-selected, and cloned into the NotI and EcoRI
compatible sites of a custom modified MCS of the
pBluescript (KS+) vector. The library was normalized in 2
rounds using conditions adapted from Soares et al., PNAS
(1994) 91: 9228-9232 and Bonaldi et al., Genome Research 6
(1996): 791, except that a significantly longer
reannealing hybridization was used."

ORIGIN

Alignment Scores:

Pred. No.: 1.01e-84 Length: 792
Score: 819.50 Matches: 164
Percent Similarity: 78.99% Conservative: 39
Best Local Similarity: 63.81% Mismatches: 53
Query Match: 59.17% Indels: 3
DB: Gaps: 1

US-10-649-273-2_copy_148_414 (1-267) x BU403563 (1-792)

```
Oy 5 AlaleuThrlleAargleuThraShlyValGluPherProPhleuValleuLeuIleSer 24
Db 2 GCATTACCACTGACAGACAGACAGATAGAAATTCCTTCTTAGTCTTACTCTCC 61
Oy 25 GlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLys 44
Db 62 GGAGGTCACTCCAT-CTGGCAGTAGCACGAGAGTTTCAAGTTCTTCTGTTGACAG 120
Oy 45 SerLeuAspIleAlaAProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIle 64
Db 121 TCCATGATATATAGACACGAGTGCATGCTGATAAGTAGACAGAAAGGCTCTTAGTG 180
Oy 65 LysHisProGluCysSerThrMetSerGlyGlyValAlaIleGlnHisLeuAlaLysGln 84
Db 181 AAGCACCAGGAGTCCACAGCATGGCCGGGAGAAAGCAATAGACACCTGGCTCAAAAC 240
Oy 85 GlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPhe 104
Db 241 GGAGACTGGCAACAGTACACTTTCAGACTTCCCATGCACAGATATCTGATATTTT 300
Oy 105 SerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLysGlnLysGln 124
Db 301 TCTTCTCCGAGCTTCAGAGCCTTGTCAAAAGCCATTCTTCAGAAAGAAAGAGAA 360
Oy 125 GlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHis 144
Db 361 GGTATTCAGAAAGGGAATCCTGTCTGCGCTTAGAGCATCGCTGCTGCACAGAC 420
Oy 145 ThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArg 164
Db 421 GTATGTGGTCTCATATATATCCAGGAGACACCGAGCCATGCTTCTGCATGAAAC 480
Oy 165 AspLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPhe 184
Db 481 AAGCATATTTATCCAAAAACGCAACTGGTGTATCAGAGAGAGTTGCATGATATCAG 540
Oy 185 TyrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysPro 204
Db 541 TATATCGAAAAGACCTGCAGACTCTGCAATGCAACCGTTTCTTCTTCTCTCT 600
Oy 205 ProProArgLeuCysThrAspAsnGlyIleMetIleAlaTPAsnGlyIleGluArgLeu 224
Db 601 CCTCAAGGCTGTGACCGATATAGTGTATGATTGATGATGATGATGATGATGATGATG 660
Oy 225 ArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrGlnProLysCysPro 244
Db 661 CGTGCAGATGTGAT-TTATACAGTACTGATGCAATCCGCTACGAACCAAAAGCTCCC 719
Oy 245 LeuGlyValAspIleSerLysGlnValGlyGlnAlaSerIleLysValPro 261
Db 720 CTTGGAATG---ATATTCCAAAGAGTTGAAGAGATCATCAAGTGCCA 767
```

RESULT 14
CN823245
LOCUS
DEFINITION
Oa.sp1bn.04N08.M13reverse.Sheep.spleen.brain.pspori1.library.Ovis
CN823245
CN823245.1 GI:47951314
EST.
SOURCE
ORGANISM
Ovis aries (sheep)
Ovis aries
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Caprinae; Ovis.
1 (bases 1 to 730)
Gosner,A. and Hopkins,J.
Ovine spleen/brain cdna library
Unpublished (2004)
COMMENT
Contact: J Hopkins

Veterinary Biomedical Sciences
University of Edinburgh
Summerhall Square, Edinburgh, EH9 1QH.
Email: j.hopkins@ed.ac.uk
Plate: 04 row: N column: 08
Seq primer: M13reverse
High quality sequence start: 6
High quality sequence stop: 550.
Location/Qualifiers
1..730
/organism="Ovis aries"
/mol_type="mRNA"
/db_xref="taxon:9940"
/clone="Oa.sp1bn.04N08"
/clone_lib="Sheep spleen/brain.pspori1.library"
/note="Vector: pspori1"

ALIGNMENT Scores:

Pred. No.:	2..3e-84	Length:	730
Score:	816.00	Matches:	157
Percent Similarity:	93.30%	Conservative:	10
Best Local Similarity:	87.71%	Mismatches:	12
Query Match:	58.92%	Indels:	0
DB:	7	Gaps:	0

US-10-649-273-2_COPY_148_414 (1-267) x CN823245 (1-730)

```
Oy 1 MetGluAlaHisAlaLeuThrIleArgLeuThraShlyValGluPherProPhleuVal 20
Db 193 ATGGAGGCTCAGACCTTACTATTAGTTAGTAAAGTAAAGTAAATTCATTATTTAGTT 252
Oy 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
Db 253 CTTTGATTTCTGAGGATATGCTTTTGCAATTAAGCTTAGAGAGATTTCAGATTTCTT 312
Oy 23 CTTTGATTTCTGAGGATATGCTTTTGCAATTAAGCTTAGAGAGATTTCAGATTTCTT 312
Db 41 LeuLeuGlyLysSerLeuAspIleAlaAProGlyAspMetLeuAspLysValAlaArgArg 60
Db 313 CTTCTTGGAAATCTTTTGGACATPAGCACAGGCGACATCTTGACAAAGTAGCAAGAG 372
Oy 61 LeuSerLeuIleLysHisProGluCysSerThrMetSerGlyGlyLysAlaIleGlnHis 80
Db 373 CTTTGCTTAATTAACATCCAGATGCTCCACCATAGATGGCGGAAAGCTATAGAACAT 432
Oy 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 100
Db 433 TTGGCCAAACAGGAAATAGATTTGATTTTCAACCTCCATGCAACGTCCTAAA 492
Oy 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120
Db 493 AATTGTGATTTTCTTTTCTGACTTCAACATGTTATGATAGATGATAGATGATGATG 552
Oy 121 GlyLysGlnGluGlyIleGlnLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaAla 140
Db 553 GAAAAGAGAGATTCAGACAGGCGCAGATCTCTTTCAGCTCAGACATTTGCTGCT 612
Oy 141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160
Db 613 GCAAGTCCAGACACCGTACCTGCCAGTGCNAAAAGAACACATGCGCTATTTCTGTT 672
Oy 161 CysLysGlnArgAspLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGly 179
Db 673 TGCACCAAGAGAGCTTTGTTTCACTCAATGAATAGCGATCTGTTGATCTGAGGC 729
```

RESULT 15
AV602901/c
LOCUS
DEFINITION
AV602901 Bos taurus kidney fetus Bos taurus cdna clone BIK1013A07
3', mRNA sequence.
AV602901
AV602901.1 GI:9725227
EST.
SOURCE
Bos taurus (cow)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2005, 15:02:11 ; Search time 64.4261 Seconds
(without alignments)
1434.756 Million cell updates/sec

Title: US-10-649-273-2_COPY_176_414
Perfect score: 1240
Sequence: 1 LLAIVQGVDFLLGKSLDI.....DISKEVGASIKVPOLKMEI 239

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_16Dec04:*

- 1: Geneseqp1980s:*
- 2: Geneseqp1990s:*
- 3: Geneseqp2000s:*
- 4: Geneseqp2001s:*
- 5: Geneseqp2002s:*
- 6: Geneseqp2003as:*
- 7: Geneseqp2003bs:*
- 8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1240	100.0	267	5	ABG96487 Novel hum
2	1240	100.0	414	5	ABG96478 Novel hum
3	1240	100.0	414	5	ABG96481 Human O-8
4	1240	100.0	414	6	ABJ26654 Human pro
5	1203	97.0	414	5	AAE29234 Human gly
6	1203	97.0	414	6	ABG71161 Novel hum
7	1203	97.0	414	6	ABU09569 Human gly
8	960	77.4	364	6	ADA54471 Human pro
9	332.5	26.8	409	4	ABG69133 Drosophil
10	326	26.3	463	3	AAV52216 Arabidops
11	319	25.7	245	3	AAI19287 Arabidops
12	319	25.7	443	3	AAI19286 Arabidops
13	319	25.7	444	3	AAI19285 Arabidops
14	257	20.7	382	6	ABU22934 Protein e
15	248.5	20.0	348	6	ABU35232 Protein e
16	247	19.9	312	6	ABU16575 Protein e
17	247	19.9	350	6	ADU05040 M. catarr
18	241	19.4	343	6	ABU39303 Protein e
19	239	19.3	251	5	AAE31054 Bhrlichia
20	231	18.6	341	4	AAU36205 Pseudomon
21	231	18.6	341	6	ABU38276 Protein e
22	231	18.6	341	7	ABU37342 P aerugin
23	231	18.6	401	7	ABO68626 Pseudomon
24	229	18.5	341	7	ADG73344 P aerugin
25	224	18.1	342	3	AAV52202 Haemophil

26	224	18.1	342	4	AAU35450 Haemophil
27	224	18.1	342	6	ABU30280 Protein e
28	213.5	17.2	347	4	ABG61112 Drosophil
29	213	17.2	337	4	AAU38187 Salmoneil
30	213	17.2	337	6	ABU47561 Protein e
31	212	17.1	340	6	ABU40514 Protein e
32	212	17.1	337	7	ADU06228 Bacteri
33	210	16.9	337	6	ABU50237 Protein e
34	209	16.9	326	4	AAE96423 Putative
35	208	16.8	335	6	ABU27480 Protein e
36	207.5	16.7	341	6	ABU40069 Protein e
37	206	16.6	337	3	AAV52204 Becherich
38	206	16.6	337	4	AAU34711 E. coli c
39	206	16.6	337	6	ABU28771 Protein e
40	205	16.5	343	7	ABO62704 Klebsiell
41	201.5	16.2	421	5	ABG96491 Novel hum
42	200	16.1	325	2	AAE26135 Glycoprot
43	200	16.1	325	3	AAV52203 Pasteurel
44	199.5	16.1	350	6	ABU25741 Protein e
45	199	16.0	354	6	ABU37844 Protein e

ALIGNMENTS

RESULT 1
ABG96487
ID ABG96487 standard; protein; 267 AA.
XX AC ABG96487;
XX DT 11-DEC-2002 (first entry)
XX DE Novel human metalloprotease MPI fragment #1.
XX
XX KW Metalloprotease; MP-1; immune disorder; glutamate transport; cancer;
XX KW motor neuron disorder; amyotrophic lateral sclerosis; ALS; diabetes;
XX KW reproductive disorder; Klinefelter's syndrome; germinal cell aplasia;
XX KW genital wart; metabolic disorder; premature puberty; Kallman syndrome;
XX KW Cushing's syndrome; neurodegenerative disease; Alzheimer's disease;
XX KW Parkinson's disease; Huntington's disease; Tourette syndrome; sepsis;
XX KW liver disease; renal disease; immune disorder; rheumatoid arthritis;
XX KW acquired immunodeficiency syndrome; AIDS; pulmonary disease; pneumonia;
XX KW emphysema; cystic fibrosis; vascular disorder; inflammatory disorder;
XX KW neurological disorder.
XX OS Homo sapiens.
XX PN WO200272751-A2.
XX PD 19-SEP-2002.
XX PF 05-FEB-2002; 2002WO-US003353.
XX PR 05-FEB-2001; 2001US-0266518P.
XX PR 10-APR-2001; 2001US-0282814P.
XX PA (BRIM) BRISTOL-MYERS SQUIBB CO.
XX PI Chen J, Feder J, Nelson TC, Duclos F, Krystek S;
XX DR MPI: 2002-723329/78.
XX DR N-PSDB; ABS76639.
XX PT New isolated nucleic acid encoding MP-1 protein, useful for preventing,
XX PT treating, or ameliorating diseases associated with aberrant
XX PT metalloproteinase activity, e.g. immune, metabolic, inflammatory and
XX PT neurological disorders.
XX PS Claim 5; Page 29; 473pp; English.
XX CC The invention describes an isolated nucleic acid molecule (I) encoding a
XX CC metalloprotease (MP-1). (I) is useful for preventing, treating, or

CC ameliorating a medical condition, particularly an immune disorder, an
 CC aberrant glutamate transport or motor neuron disorder, such as
 CC amyotrophic lateral sclerosis (ALS), its juvenile form or an ALS-like
 CC condition. The compositions and methods are also useful for diagnosing,
 CC prognosticating, treating, ameliorating and/or treating disorders
 CC associated with MP-1 activity, e.g. diabetes, cancer, reproductive
 CC disorders (e.g. Klinefelter's syndrome, genital warts, or germinal cell
 CC aplasia), metabolic disorders (e.g. premature puberty, Kallman syndrome,
 CC or Cushing's syndrome), neurodegenerative diseases (Alzheimer's disease,
 CC Parkinson's disease, Huntington's disease or Tourette syndrome), liver
 CC and renal diseases and immune disorders (e.g. AIDS, rheumatoid arthritis
 CC or sepsis), pulmonary diseases (e.g. pneumonia, emphysema or cystic
 CC fibrosis) and vascular, inflammatory and neurological disorders (e.g.
 CC Alzheimer's disease or Parkinson's disease). This is the amino acid
 CC sequence of a metalloprotease MP1 protein

SQ Sequence 267 AA;

Query Match 100.0%; Score 1240; DB 5; Length 267;
 Best Local Similarity 100.0%; Pred. No. 7.9e-131;
 Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LIALVQGVSDPFLILGKSLDIAPGDMLDKVARRLSLIKHPECSTMGSKAIEHLAKQGNRF 60
 DB 29 LIALVQGVSDPFLILGKSLDIAPGDMLDKVARRLSLIKHPECSTMGSKAIEHLAKQGNRF 88
 QY 61 HFDIKRPLHAKNCDPSFTGLQHTVDKIIIMKKEKEGIEKGQILSSAADIAATVQHTMAC 120
 DB 89 HFDIKRPLHAKNCDPSFTGLQHTVDKIIIMKKEKEGIEKGQILSSAADIAATVQHTMAC 148
 QY 121 HLVKRTRRALILFCRKQDILPQNNAVLVASGVASNFYIRALIEILTNATQCTLLCPPEPL 180
 DB 149 HLVKRTRRALILFCRKQDILPQNNAVLVASGVASNFYIRALIEILTNATQCTLLCPPEPL 208
 QY 181 CTGNGIMIAMNGIERLRAGILGILHDIEGIRYEPKCPGLVDISKVEGASIKVPOLKMEI 239
 DB 209 CTGNGIMIAMNGIERLRAGILGILHDIEGIRYEPKCPGLVDISKVEGASIKVPOLKMEI 267

RESULT 2
 ABG96478
 ID ABG96478 standard; protein; 414 AA.
 AC ABG96478;
 XX
 XX 11-DEC-2002 (first entry)
 DT
 XX
 XX Novel human metalloprotease MP1.
 DE
 XX Metalloprotease; MP-1; immune disorder; glutamate transport; cancer;
 KW motor neuron disorder; amyotrophic lateral sclerosis; ALS; diabetes;
 KW reproductive disorder; Klinefelter's syndrome; germinal cell aplasia;
 KW genital wart; metabolic disorder; premature puberty; Kallman syndrome;
 KW Cushing's syndrome; neurodegenerative disease; Alzheimer's disease;
 KW Parkinson's disease; Huntington's disease; Tourette syndrome; sepsis;
 KW liver disease; renal disease; immune disorder; rheumatoid arthritis;
 KW acquired immunodeficiency syndrome; AIDS; pulmonary disease; pneumonia;
 KW emphysema; cystic fibrosis; vascular disorder; inflammatory disorder;
 KW neurological disorder.
 KW
 OS Homo sapiens.
 XX
 XX WO200272751-A2.
 FN
 XX
 XX 19-SEP-2002.
 PD
 XX
 XX 05-FEB-2002; 2002MO-US003353.
 PF
 XX
 XX 05-FEB-2001; 2001US-0266518P.
 PR
 XX 10-APR-2001; 2001US-0282814P.
 XX
 XX (BRIM) BRISTOL-MYERS SQUIBB CO.

PI Chen J, Feder J, Nelson TC, Duclos F, Krystek S;
 XX
 XX MPI; 2002-72329/78.
 DR N-PSDB; ABS76635.
 XX
 PT New isolated nucleic acid encoding MP-1 protein, useful for preventing,
 PT treating, or ameliorating diseases associated with aberrant
 PT metalloprotease activity, e.g. immune, metabolic, inflammatory and
 PT neurological disorders.
 XX
 PS Claim 5; Fig 1A-C; 473pp; English.

CC The invention describes an isolated nucleic acid molecule (I) encoding a
 CC metalloprotease (MP-1). (I) is useful for preventing, treating, or
 CC ameliorating a medical condition, particularly an immune disorder, an
 CC aberrant glutamate transport or motor neuron disorder, such as
 CC amyotrophic lateral sclerosis (ALS), its juvenile form or an ALS-like
 CC condition. The compositions and methods are also useful for diagnosing,
 CC prognosticating, treating, ameliorating and/or treating disorders
 CC associated with MP-1 activity, e.g. diabetes, cancer, reproductive
 CC disorders (e.g. Klinefelter's syndrome, genital warts, or germinal cell
 CC aplasia), metabolic disorders (e.g. premature puberty, Kallman syndrome,
 CC or Cushing's syndrome), neurodegenerative diseases (Alzheimer's disease,
 CC Parkinson's disease, Huntington's disease or Tourette syndrome), liver
 CC and renal diseases and immune disorders (e.g. AIDS, rheumatoid arthritis
 CC or sepsis), pulmonary diseases (e.g. pneumonia, emphysema or cystic
 CC fibrosis) and vascular, inflammatory and neurological disorders (e.g.
 CC Alzheimer's disease or Parkinson's disease). This is the amino acid
 CC sequence of a metalloprotease MP1 protein

SQ Sequence 414 AA;

Query Match 100.0%; Score 1240; DB 5; Length 414;
 Best Local Similarity 100.0%; Pred. No. 1.5e-130;
 Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LIALVQGVSDPFLILGKSLDIAPGDMLDKVARRLSLIKHPECSTMGSKAIEHLAKQGNRF 60
 DB 176 LIALVQGVSDPFLILGKSLDIAPGDMLDKVARRLSLIKHPECSTMGSKAIEHLAKQGNRF 235
 QY 61 HFDIKRPLHAKNCDPSFTGLQHTVDKIIIMKKEKEGIEKGQILSSAADIAATVQHTMAC 120
 DB 236 HFDIKRPLHAKNCDPSFTGLQHTVDKIIIMKKEKEGIEKGQILSSAADIAATVQHTMAC 295
 QY 121 HLVKRTRRALILFCRKQDILPQNNAVLVASGVASNFYIRALIEILTNATQCTLLCPPEPL 180
 DB 296 HLVKRTRRALILFCRKQDILPQNNAVLVASGVASNFYIRALIEILTNATQCTLLCPPEPL 355
 QY 181 CTGNGIMIAMNGIERLRAGILGILHDIEGIRYEPKCPGLVDISKVEGASIKVPOLKMEI 239
 DB 356 CTGNGIMIAMNGIERLRAGILGILHDIEGIRYEPKCPGLVDISKVEGASIKVPOLKMEI 414

RESULT 3
 ABB05481
 ID ABB05481 standard; protein; 414 AA.
 AC ABB05481;
 XX
 XX 19-APR-2002 (first entry)
 DT
 XX
 XX Human O-sialoglycoproteinase-like protein SEQ ID NO:2.
 DE Human, O-sialoglycoproteinase-like protein; OSGPLP, enzyme.
 KW Human, O-sialoglycoproteinase-like protein; OSGPLP, enzyme.
 XX
 XX Homo sapiens.
 OS
 XX
 XX CN1318550-A.
 PN
 XX
 XX 24-OCT-2001.
 PD
 XX
 XX 19-APR-2000; 2000CN-00106834.
 PF
 XX

PR 19-APR-2000; 2000CN-00106834.
XX
PA (SHAN-) SHANGHAI SHENGYUAN GENE DEV CO LTD.
PI Mao Y, Xie Y;
XX
DR WPI: 2002-115090/16.
XX N-PSDB; ABA93268.
PT O-sialoglycoproteinase-like protein and encoding polynucleotide, useful
XX for diagnosing, preventing and treating related diseases.
PS Claim 1; Page 30-31 (Disclosure); 38pp; Chinese.
XX
XX The present sequence represents human O-sialoglycoproteinase-like protein
CC (OSGPLP). The present invention also describes: (1) the preparation of
CC the OSGPLP protein; (2) applying the OSGPLP protein in diagnosis; (3) the
CC prevention and/or treatment of related diseases; (4) utilizing the OSGPLP
CC protein in screening its agonist, excitomotor and inhibitor and preparing
CC an antibody against the OSGPLP protein; and (5) the use of the OSGPLP
CC polynucleotide sequences, proteins, agonists, excitomotors, inhibitors
CC and antibodies in treating diseases related to the abnormal OSGPLP gene
CC and in preparing the medicine composite for the treatment
XX
SQ Sequence 414 AA;
Query Match 100.0%; Score 1240; DB 5; Length 414;
Best Local Similarity 100.0%; Pred. No. 1.5e-130;
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 LIALVGVSDPFLILGKSLDIAPGDMLDKVARRLSLIKHEPCSTMSGKAIIEHLAKGNRF 60
DB 176 LIALVGVSDPFLILGKSLDIAPGDMLDKVARRLSLIKHEPCSTMSGKAIIEHLAKGNRF 235
QY 61 HFDIKPPLHAKNCDFSTGLQHTVDKTIIMKEKEKEGIEKGQILSSAADIAATVQHTMAC 120
DB 236 HFDIKPPLHAKNCDFSTGLQHTVDKTIIMKEKEKEGIEKGQILSSAADIAATVQHTMAC 295
QY 121 HLVKRTTRAILFCQKRDLLPQNNNAVLVASGVASNFYIRRALEILTNAQCTLLCPPPRL 180
DB 296 HLVKRTTRAILFCQKRDLLPQNNNAVLVASGVASNFYIRRALEILTNAQCTLLCPPPRL 355
QY 181 CTDNGIMIANWGIERLRAGLILHDIEGIRYEPKCPGLVDISKVEGASIKVPOLKMEI 239
DB 356 CTDNGIMIANWGIERLRAGLILHDIEGIRYEPKCPGLVDISKVEGASIKVPOLKMEI 414
RESULT 4
ABJ26654
ID ABJ26654 standard; protein; 414 AA.
XX
AC ABJ26654;
XX
DT 01-MAY-2003 (first entry)
XX
DE Human protein modification + maintenance molecule protein SEQ ID No 8.
XX
XX Cytostatic; antiarteriosclerotic; anti-HIV; antiallergic; anticonvulsant;
XX cerebroprotective; antiparkinsonian; nootropic; antiinflammatory;
XX antitumor; hepatotropic; gynaecological; antibacterial; virucide;
XX protooncogene; antiparasitic; cell proliferative disease; PMOD;
XX protein modification and maintenance molecule; immunogenic fragment;
XX cancer; autoimmune; inflammatory disease; neurological disorder;
XX gastrointestinal; developmental; vesicle trafficking disorder; infection;
XX protein-protein interaction; drug-target interaction;
XX gene expression profile; human.
XX
XX Homo sapiens.
XX
XX WC2003000944-A2.
XX
XX 03-JAN-2003.
XX

PF 18-JUN-2002; 2002WO-US019360.
XX
XX 22-JUN-2001; 2001US-0300508P.
XX
XX 06-JUL-2001; 2001US-0303445P.
XX
XX 13-JUL-2001; 2001US-0305405P.
XX
XX 09-AUG-2001; 2001US-0311442P.
XX
XX 24-AUG-2001; 2001US-0314621P.
XX
XX 29-AUG-2001; 2001US-0315992P.
XX
XX 03-MAY-2002; 2002US-0378205P.
PA (INCY-) INCYTE GENOMICS INC.
XX
XX Gandhi AR, Kable AE, Swarnakar A, Hafalia ALA, Tran B, Duggan BM,
XX Warren BA, Ison CH, Honchell CD, Nguyen DB, Lu DAM, Lee EA, Yue H;
XX Forrester J, Barroso I, Ramkumar J, Griffin JA, Li JX, Yang J;
XX Thangavelu K, Gietzen KJ, Ding L, Baughn MR, Boroweky ML, Yao MG;
XX Walla NK, Mason PM, Gunturjan R, Lee S, Becha SD, Lee SY, Tran UK;
XX Elliott VS, Luo W, Sprague WW, Tang YT, Lu Y, Zebardian Y;
XX
XX WPI: 2003-184039/18.
XX N-PSDB; ABR23207.
PT New isolated human PMOD polypeptide and polynucleotide, useful for
PT diagnosing, treating and preventing diseases or conditions associated
PT with the aberrant PMOD expression, e.g. cancer, AIDS, atherosclerosis and
PT infections.
XX
PS Claim 63; Page 182-183; 225pp; English.
XX
XX The invention relates to an isolated polypeptide comprising: any of 28
XX sequences of 48-1256 amino acids; a natural amino acid sequence at least
XX 90% identical to the 28 amino acid sequences, 94% identical to a sequence
XX of 703 or 267 amino acids, 96% identical to a sequence of 414 amino
XX acids, or 97% identical to a sequence of 242 amino acids, all given in
XX the specification; or a biologically active or immunogenic fragment of
XX the isolated polypeptide. The polypeptides and polynucleotides are useful
XX in diagnosing, treating and preventing diseases or conditions associated
XX with the decreased expression of protein modification and maintenance
XX molecules (PMOD), such as cell proliferative diseases (e.g. cancer,
XX atherosclerosis), autoimmune/inflammatory diseases (e.g. AIDS,
XX allergies), neurological disorders (e.g. stroke, Parkinson's disease,
XX epilepsy), gastrointestinal (e.g. ulcer, cirrhosis), reproductive (e.g.
XX endometriosis), developmental, vesicle trafficking disorders, and
XX infections (e.g. bacterial, viral, parasitic, protozoal). These are also
XX useful in assessing the effects of exogenous compounds on the expression
XX of nucleic acid and amino acid sequences of PMOD. The PMOD or its
XX fragments are useful in screening compounds for effectiveness as agonist
XX or antagonist of the polypeptides, or in altering the expression of the
XX target polynucleotide and compounds that specifically bind to or modulate
XX the activity of the polypeptide. The microarray is useful in monitoring
XX or measuring protein-protein interactions, drug-target interactions, and
XX gene expression profiles. This sequence represents a human PMOD protein
XX of the invention
XX
SQ Sequence 414 AA;
Query Match 100.0%; Score 1240; DB 6; Length 414;
Best Local Similarity 100.0%; Pred. No. 1.5e-130;
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 LIALVGVSDPFLILGKSLDIAPGDMLDKVARRLSLIKHEPCSTMSGKAIIEHLAKGNRF 60
DB 176 LIALVGVSDPFLILGKSLDIAPGDMLDKVARRLSLIKHEPCSTMSGKAIIEHLAKGNRF 235
QY 61 HFDIKPPLHAKNCDFSTGLQHTVDKTIIMKEKEKEGIEKGQILSSAADIAATVQHTMAC 120
DB 236 HFDIKPPLHAKNCDFSTGLQHTVDKTIIMKEKEKEGIEKGQILSSAADIAATVQHTMAC 295
QY 121 HLVKRTTRAILFCQKRDLLPQNNNAVLVASGVASNFYIRRALEILTNAQCTLLCPPPRL 180
DB 296 HLVKRTTRAILFCQKRDLLPQNNNAVLVASGVASNFYIRRALEILTNAQCTLLCPPPRL 355
QY 181 CTDNGIMIANWGIERLRAGLILHDIEGIRYEPKCPGLVDISKVEGASIKVPOLKMEI 239

```

Db          356 CTDNGIMIAMNGIERLAGILHDIEGIRYBPKCPGLVDISKVEGASIKVPOLKMEI 414
              |||
RESULT 5
AAE29234
XX AAE29234 standard; protein; 414 AA.
XX
XX AAE29234;
XX
XX 27-JAN-2003 (first entry)
XX
XX Human glycoprotease 28472 protein.
XX
XX Human; adenosine deaminase; seven transmembrane domain receptor; cancer;
XX 7TM; glycoprotease; immune disorder; Iga deficiency; allergy; arrhythmia;
XX rheumatoid arthritis; diabetes; atherosclerosis; cardiovascular disorder;
XX hypertension; ischaemic heart disease; obesity; myocardial infarction;
XX endothelial cell disorder; Grave's disease; psoriasis; brain disorder;
XX Parkinson's disease; Alzheimer's disease; haematopoietic disorder;
XX cerebral oedema; metabolic disorder; liver disorder; platelet disorder;
XX chromosome mapping; tissue typing; gene therapy; neuroprotective;
XX cytoskeletal; anorectic; cardiac; haemostatic.
XX
XX Homo sapiens.
XX
FH Key      Location/Qualifiers
FT Domain  1..108
FT          /note= "Non-transmembrane domain; N-terminal cytoplasmic
FT          domain"
FT          38..369
FT          /note= "Endopeptidase O-sialoglycoprotein hydrolase
FT          metalloprotease zinc glycoprotease sialoglycoprotease
FT          domain"
FT          109..132
FT          /note= "Transmembrane domain"
FT          133..164
FT          /note= "Non-transmembrane domain; non-cytoplasmic loop"
FT          138..152
FT          /note= "Glycoprotease domain"
FT          165..189
FT          /note= "Transmembrane domain"
FT          190..316
FT          /note= "Non-transmembrane domain; cytoplasmic domain"
FT          317..333
FT          /note= "Transmembrane domain"
FT          334..414
FT          /note= "Non-transmembrane domain"
FT          374..414
FT          /note= "Sialoglycoprotease type domain"
FT Domain
FT
XX WO200274960-A2.
XX
XX 26-SEP-2002.
XX
XX 08-NOV-2001; 2001WO-US051427.
XX
XX 08-NOV-2000; 2000US-0246768P.
XX 08-NOV-2000; 2000US-0246772P.
XX 15-NOV-2000; 2000US-0249185P.
XX
XX (MILL-) MILLENNIUM PHARM INC.
XX
XX Leiby KR, Kapeller-Libermann R, Glucksmann M;
XX
XX WPI; 2002-759898/82.
XX
XX N-PSDB; AAD46856.
XX
XX New 38650, 28472, 5495, 65507, 81588 or 14354 nucleic acid molecules,
XX useful for diagnosing and treating cancer, immune, cardiovascular,
XX hematoopoietic, brain, pain, metabolic, liver or platelet disorders, and
XX in pharmacogenomics.

```

```

PS Claim 1; Fig 8; 178pp; English.
XX
XX The present invention relates to novel 38650, 28472, 5495, 65507, 81588
XX or 14354 nucleic acid molecules which encode adenosine deaminase, glyco-
XX protease or seven transmembrane domain (7TM) receptor family members.
XX Sequences of the invention are useful in diagnosing and treating cancer
XX or aberrant cellular proliferation and/or differentiation (e.g. colon or
XX lung cancer), immune disorders (e.g. selective Iga deficiency, rheumatoid
XX arthritis, diabetes, allergy), heart or cardiovascular disorders (e.g.,
XX hypertension, atherosclerosis, arrhythmias, ischaemic heart disease,
XX myocardial infarction, thrombus) including endothelial cell disorders
XX (e.g. psoriasis, Grave's disease), haematopoietic disorders, brain
XX disorders (e.g. cerebral oedema, Parkinson's or Alzheimer's disease),
XX pain and metabolic disorders (e.g. obesity), liver disorders or platelet
XX disorders. They are also useful in screening assays, predictive medicine
XX (e.g. diagnostic assays, prognostic assays, monitoring clinical trials
XX and pharmacogenetics) and prophylactic and therapeutic methods. The
XX nucleic acids may also be used in chromosome mapping, tissue typing and
XX forensic biology and as surrogate markers. Sequences of the invention are
XX also used in gene therapy. The present sequence is human glycoprotease
XX 28472 protein
XX
SQ Sequence 414 AA;
XX
XX Query Match      97.0%; Score 1203; DB 5; Length 414;
XX Best Local Similarity 97.1%; Pred. No. 2,3e-126;
XX Matches 232; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
XX
QY 1 LIALVQGVSDFTLLGKSLDIAPGDMIDKVARRLSLIKHEPCSTMSGKAIIEHLAKQGNR 60
XX |||
DB 176 LIALVQGVSDFTLLGKSLDIAPGDMIDKVARRLSLIKHEPCSTMSGKAIIEHLAKQGNR 235
XX |||
QY 61 HPDIKPEPLHANNCDSFTGLQHVTDKIMKKEKKEKIGKQILSSAADIATVQHTMAC 120
XX |||
DB 236 HPDIKPEPLHANNCDSFTGLQHVTDKIMKKEKKEKIGKQILSSAADIATVQHTMAC 235
XX |||
QY 121 HLKRTTRAILPCKQKDLIPONNAVVAAGVANSFYIRRALEIILTNATQCTLLCPPPRL 180
XX |||
DB 296 HLKRTTRAILPCKQKDLIPONNAVVAAGVANSFYIRRALEIILTNATQCTLLCPPPRL 355
XX |||
QY 181 CTDNGIMIAMNGIERLAGILHDIEGIRYBPKCPGLVDISKVEGASIKVPOLKMEI 239
XX |||
DB 356 CTDNGIMIAMNGIERLAGILHDIEGIRYBPKCPGLVDISKVEGASIKVPOLKMEI 414
XX |||
XX
XX RESULT 6
XX ABG71161
XX ID ABG71161 standard; protein; 414 AA.
XX
XX AC ABG71161;
XX
XX 30-JAN-2003 (first entry)
XX
XX Novel human glycoprotease 28472.
XX
XX Cancer; aberrant cell proliferation; aberrant cell differentiation;
XX breast cancer; ovarian cancer; prostate cancer; colon cancer;
XX lung cancer; immune disorder; heart disorder; cardiovascular disorder;
XX endothelial disorder; hematopoietic disorder; blood vessel disorder;
XX brain disorder; pain; metabolic disorder; liver disorder; diabetes;
XX platelet disorder; carcinoma; sarcoma; leukaemia; Hodgkin's disease;
XX autoimmune disorder; hypertension; atherosclerosis; heart failure;
XX myocardial infarction; ischaemic heart disease; Crohn's disease;
XX Grave's disease; Kawasaki syndrome; Raynaud's disease; aneurysm;
XX cerebral ischaemia; peripheral neuropathy; Alzheimer's disease;
XX Parkinson's disease; anorexia nervosa; cachexia; glycoprotease; enzyme.
XX
XX Homo sapiens.
XX
XX Key      Location/Qualifiers
XX Domain  138..152
XX          /label= Glycoprotease_domain
XX

```


PN WO00277233-A2.
XX
XX 03-OCT-2002.
XX
XX 08-NOV-2001; 2001WO-US046724.
XX
XX 08-NOV-2000; 2000US-0246768P.
PR 08-NOV-2000; 2000US-0246772P.
PR 15-NOV-2000; 2000US-0249185P.
XX
XX (MILL-) MILLENNIUM PHARM INC.
PI Leiby KR, Kapeller-Libermann R, Glucksmann M;
XX WPI; 2003-029938/02.
DR N-PSDB; ABSS57020.
XX
XX New adenosine deaminase, glycoprotease and seven transmembrane domain
PT nucleic acids and polypeptides, designated 38650, 28472, 5495, 65507,
PT 81588 and 14354, useful for treating e.g. leukemia, Hodgkin's disease or
PT hypertension.
XX
XX Claim 4; Fig 8A-B; 178pp; English.
XX
XX The invention describes isolated 38650, 28472, 5495, 65507, 81588 and
CC 14354 nucleic acid molecules (I) and their encoded polypeptides (II). The
CC 38650 nucleic acid molecule comprises a sequence encoding adenosine
CC deaminase. The 28472 nucleic acid molecule comprises a sequence encoding
CC a human glycoprotease. The 5495, 65507, 81588 and 14354 comprise
CC sequences that encode a human seven transmembrane domain (7TM). The
CC 38650, 28472, 5495, 65507, 81588 or 14354 nucleic acid and polypeptide
CC sequences are useful for diagnosing, preventing or treating a subject
CC with or at risk of developing a disorder, e.g. cancer or aberrant
CC cellular proliferation and/or differentiation (e.g. breast, ovarian,
CC prostate, colon or lung cancer), immune disorders, heart disorders,
CC cardiovascular disorders, endometrial disorders, hematopoietic disorders,
CC blood vessel disorders, brain disorders, pain and metabolic disorders,
CC liver disorders or platelet disorders. These disorders include carcinoma,
CC sarcoma, leukemia, Hodgkin's disease, autoimmune disorders,
CC hypertension, atherosclerosis, heart failure, myocardial infarction,
CC ischemic heart disease, Crohn's disease, Grave's disease, Kawasaki
CC syndrome, Raynaud's disease, aneurysm, cerebral ischemia, peripheral
CC neuropathy, Alzheimer's disease, Parkinson's disease, anorexia nervosa,
CC cecilia or diabetes. This is the amino acid sequence of the novel human
CC glycoprotease 28472
XX
XX Sequence 414 AA;
SQ
Query Match 97.0%; Score 1203; DB 6; Length 414;
Best Local Similarity 97.1%; Pred. No. 2,3e-126;
Matches 232; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
QY 1 LTAIVGVSDPFLKSLDIAPGDMLDKVARRLSLIKHPCSTMSGKRAIHLAKQGNRF 60
DB 176 LTAIVGVSDPFLKSLDIAPGDMLDKVARRLSLIKHPCSTMSGKRAIHLAKQGNRF 235
QY 61 HFDIKRPLHAKNCDFSTFGQHTVDKTIKMKKEKEGIEKGQILSSAADIAATVQHTMAC 120
DB 236 HFDIKRPLHAKNCDFSTFGQHTVDKTIKMKKEKEGIEKGQILSSAADIAATVQHTMAC 295
QY 121 HLVTTRRAILFCORBDLPPONNAVIVASGVASFYRRALEIITNATQCTLCPPL 180
DB 296 HLVTTRRAILFCORBDLPPONNAVIVASGVASFYRRALEIITNATQCTLCPPL 355
QY 181 CTDNGMIAMNGIERLRAGILHDIEGIRYRKPCLGVDSKEVGEASIKVPOLKMEI 239
DB 356 CTDNGMIAMNGIERLRAGILHDIEGIRYRKPCLGVDSKEVGEASIKVPOLKMEI 414

AC ABU09569;
XX
XX 08-JUL-2003 (first entry)
XX
XX Human glycoprotease encoded by cDNA 28472.
XX
XX Human, enzyme; cancer; aberrant cellular proliferation; differentiation;
XX immune disorders; heart disorder; brain disorder;
XX cardiovascular disorder; endometrial cell disorder; pain disorder;
XX haematopoietic disorder; blood vessel disorder; metabolic disorder;
XX liver disorder; platelet disorder; glycoprotease.
XX
XX Homo sapiens.
XX
XX US2003009017-A1.
XX
XX 09-JUN-2003.
XX
XX 08-NOV-2001; 2001US-00012140.
XX
XX 08-NOV-2000; 2000US-0246768P.
PR 08-NOV-2000; 2000US-0246772P.
PR 15-NOV-2000; 2000US-0249185P.
XX
XX (LEIB/) LEIBY K R.
XX (KAP/) KAPPELLER-LIBERMANN R.
XX (GLUC/) GLUCKSMANN M A.
PI Leiby KR, Kapeller-Libermann R, Glucksmann MA;
XX WPI; 2003-428888/40.
DR N-PSDB; ACA60887.
XX
XX New isolated 38650, 28472, 5495, 65507, 81588 and 14354 nucleic acid
PT molecules, useful for diagnosing, treating cancer, pain, or immune,
PT heart, endometrial cell, hematopoietic, blood vessel, brain, metabolic
PT and liver disorders.
XX
XX Claim 4; Fig 8; 90pp; English.
XX
XX The invention relates to an isolated 38650 (encoding adenosine
CC deaminase), 28472 (encoding a glycoprotease), 5495 (encoding a 7
CC transmembrane receptor), 65507, 81588 or 14354 nucleic acid molecule (or
CC a sequence which is at least 60% identical to the six nucleic acids or
CC their open reading frames, fragments of at least 15 nucleotides,
CC naturally occurring variants, or a DNA insert of the plasmid deposited
CC with the American Type Culture Collection as Accession No. not defined in
CC the specification, which encodes the amino acid sequence). Also included
CC are a host cell containing the nucleic acids (used to produce the
CC proteins), the encoded proteins, an antibody that selectively binds to
CC the polypeptide, and identifying a compound that binds to/modulates the
CC activity of the polypeptide. The nucleic acid molecules, polypeptides and
CC methods are useful for diagnosing, treating cancer, aberrant cellular
CC proliferation and/or differentiation, immune disorders, heart disorders,
CC cardiovascular disorders including endometrial cell disorders,
CC haematopoietic disorders, blood vessel disorders, brain disorders, pain
CC and metabolic disorders, liver disorders and platelet disorders (many
CC examples of these disorders are given in the specification). The present
CC sequence is the Human glycoprotease encoded by cDNA 28472
XX
XX Sequence 414 AA;
SQ
Query Match 97.0%; Score 1203; DB 6; Length 414;
Best Local Similarity 97.1%; Pred. No. 2,3e-126;
Matches 232; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
QY 1 LTAIVGVSDPFLKSLDIAPGDMLDKVARRLSLIKHPCSTMSGKRAIHLAKQGNRF 60
DB 176 LTAIVGVSDPFLKSLDIAPGDMLDKVARRLSLIKHPCSTMSGKRAIHLAKQGNRF 235
QY 61 HFDIKRPLHAKNCDFSTFGQHTVDKTIKMKKEKEGIEKGQILSSAADIAATVQHTMAC 120
DB 236 HFDIKRPLHAKNCDFSTFGQHTVDKTIKMKKEKEGIEKGQILSSAADIAATVQHTMAC 295

Qy 121 HLVRKTRRAILFCQKORDLLPQNNNAVLVASGGVASFYIRRALEILTNATQCTLLCPPRL 180
| | | | |
Db 296 HLVRKTRRAILFCQKORDLLPQNNNAVLVASGGVASFYIRRALEILTNATQCTLLCPPRL 355
| | | | |
Qy 181 CTDNGIMTANNGIERLRAGLGLIHDIBGIRYBPCKPLGVDSKEVGEASIKVPOLKMEI 239
| | | | |
Db 356 CTDNGIMTANNGIERLRAGLGLIHDIBGIRYBPCKPLGVDSKEVGEASIKVPOLKMEI 414
| | | | |

RESULT 8

ADAS4471
ID ADAS4471 standard; protein; 364 AA.

AC ADAS4471;

DT 20-NOV-2003 (first entry)

DE Human protein, SEQ ID 2039.

XX Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Noctropic;

KW Gene therapy; human; secretory protein; membrane proteins; cancer;

KM Inflammatory disease; osteoporosis; neurological disease.

XX Homo sapiens.

PN EPI293569-A2.

PD 19-MAR-2003.

PF 21-MAR-2002; 2002EP-00006586.

PR 14-SEP-2001; 2001JP-00328381.

PR 24-JAN-2002; 2002US-0350435P.

XX (HELI-) HELIX RES INST.

PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;

PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;

PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuno Y;

DR WPI; 2003-395539/38.

DR N-PSDB; ADA52832.

XX New polynucleotides encoding full-length polypeptides, e.g. secretory

PT and/or membrane proteins, useful for developing medicines for diseases in

PT which the gene is involved, or as target molecules for gene therapy.

XX Claim 14; SEQ ID NO 2039; 205pp; English.

XX The present invention relates to novel human secretory or membrane

CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-

CC ADA54071). The coding sequences are useful in the gene therapy of

CC diseases caused by abnormalities of the proteins, e.g. cancer,

CC inflammatory diseases, osteoporosis or neurological disease.

XX Sequence 364 AA;

Query Match 77.4%; Score 960; DB 6; Length 364;

Best Local Similarity 98.4%; Pred. No. 5e-99;

Matches 186; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 296 HLVRKTRRAILFCQKORDLLPQNNNAVLVASGGVASFYIRRALEILTNATQCTLLCPPRL 355
| | | | |
Qy 181 CTDNGIMTANNGIERLRAGLGLIHDIBGIRYBPCKPLGVDSKEVGEASIKVPOLKMEI 239
| | | | |
Db 356 CTDNGIMTANNGIERLRAGLGLIHDIBGIRYBPCKPLGVDSKEVGEASIKVPOLKMEI 414
| | | | |

RESULT 9

ABB69133
ID ABB69133 standard; protein; 409 AA.

AC ABB69133;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 34191.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US009231.

PR 23-MAR-2000; 2000US-0191637P.

PR 11-JUL-2000; 2000US-00614150.

XX (PEKE) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

PI WPI; 2001-656860/75.

DR N-PSDB; ABL13236.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signaling and cell-cell

PT interactions.

XX Disclosure; SEQ ID NO 34191; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-

CC ABB72072). The sequence data for this patent did not form part of the

CC printed specification, but was obtained in electronic format directly

CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 409 AA;

Query Match 26.8%; Score 332.5; DB 4; Length 409;

Best Local Similarity 35.9%; Pred. No. 3.9e-28;

Matches 83; Conservative 38; Mismatches 97; Indels 13; Gaps 5;

Db 296 HLVRKTRRAILFCQKORDLLPQNNNAVLVASGGVASFYIRRALEILTNATQCTLLCPPRL 355
| | | | |
Qy 181 CTDNGIMTANNGIERLRAGLGLIHDIBGIRYBPCKPLGVDSKEVGEASIKVPOLKMEI 239
| | | | |
Db 356 CTDNGIMTANNGIERLRAGLGLIHDIBGIRYBPCKPLGVDSKEVGEASIKVPOLKMEI 414
| | | | |

RESULT 9

ABB69133
ID ABB69133 standard; protein; 409 AA.

AC ABB69133;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 34191.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US009231.

PR 23-MAR-2000; 2000US-0191637P.

PR 11-JUL-2000; 2000US-00614150.

XX (PEKE) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

PI WPI; 2001-656860/75.

DR N-PSDB; ABL13236.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signaling and cell-cell

PT interactions.

XX Disclosure; SEQ ID NO 34191; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-

CC ABB72072). The sequence data for this patent did not form part of the

CC printed specification, but was obtained in electronic format directly

CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 409 AA;

Query Match 26.8%; Score 332.5; DB 4; Length 409;

Best Local Similarity 35.9%; Pred. No. 3.9e-28;

Matches 83; Conservative 38; Mismatches 97; Indels 13; Gaps 5;

QY 179 RLCTDNGIMIAMNGIERLRAGILHDIE-GIRYEPKCPGLVDISKVEGEA 228
DB 347 RYCSNDGVMWMTAMHGEVL-----LQDKEASTRYDYD---SIDIQSAGFA 388
RESULT 10
AAV52216
ID AAV52216 standard; protein; 463 AA.
XX AAV52216;
XX 09-FEB-2000 (first entry)
XX Arabidopsis thaliana ysjd protein homologue.
XX ysjd protein; essential; Gram positive; Gram negative; conserved; motif;
KM identification; antagonist; antibacterial; antibiotic; broad spectrum;
XX treatment; infection; resistance; drug target.
XX Arabidopsis thaliana.
XX
XX Key Location/Qualifiers
FH 86..96
FT Region /note="ysjd conserved motif 3"
FT 111..131
FT Region /note="ysjd conserved motif 4"
FT 152..198
FT Region /note="ysjd conserved motif 2"
FT 208..259
FT Region /note="ysjd conserved motif 1"
XX
XX W09954470-A2.
XX
XX 28-OCT-1999.
XX
XX 20-APR-1999; 99WO-EP002635.
XX
XX 22-APR-1998; 98GB-00008423.
XX
XX (GLAXO) GLAXO GROUP LTD.
XX
XX Arigoni F, Edgerton MD, Loferer H, Peitsch MC;
XX WPI; 2000-013253/01.
XX
XX Novel bacterial polypeptides used to identify broad spectrum antibiotics.
XX
XX Claim 1; Fig 1; 55pp; English.
XX
XX Proteins AAV52202-Y52217 encompass a novel family of proteins designated
CC the ysjd family, after the name given to the Escherichia coli family
CC member. These proteins are essential for the survival of both Gram
CC negative and Gram positive bacteria, although no function has as yet been
CC ascribed to these proteins. The ysjd proteins, fragments of ysjd proteins
CC (for example, fragments encompassing one or more conserved ysjd motifs
CC such as AAV52218-Y52284) and nucleotides encoding them can be used to
CC identify antagonists and broad spectrum antibacterial compounds. These
CC antagonists and compounds can be used to treat a wide range of bacterial
CC infections. New antibiotics are urgently needed, as serious bacterial
CC infections and antibiotic resistant strains are becoming increasingly
CC prevalent. The proteins of the invention are essential proteins for
CC bacterial viability, and represent new targets for antibiotics
XX
XX Sequence 463 AA;
SQ
Query Match 26.3%; Score 326; DB 3; Length 463;
Best Local Similarity 33.6%; Pred. No. 2.6e-27;
Matches 82; Conservative 34; Mismatches 84; Indels 44; Gaps 5;

QY 61 HFDIKRPLHAKNCPSPFGLQHWTDKTIIMKKEKEGIEKGQILSSNADIATVQHTMAC 120
DB 276 SVKFNVPKMYCHDQCNFSYAGLKYVRLAIEAKE-----IRNRADIAASFQVAVL 325
QY 121 HLVKRTRRAILPCKQBDLLPQNNAVLVASGVASNFYIRRALEILTNACQTLCPPPRL 180
DB 326 HLEEKERAILDWALE--LEPSIKHNVISGVASNKYVRLRLNIVENKRLKLVCPSPSL 382
QY 181 CTDNGIMIAMNGIERLRAGILHDIEGIRYE-----PKCPLGVDSKE 224
DB 383 CTDNGVMVAMTGLEHFRVG-----RYDPPPATRPEDVYVYDLRRPRWPLGEERYAKG 432
QY 225 VGEA 228
DB 433 RSEA 436
RESULT 11
AAG19287
ID AAG19287 standard; protein; 245 AA.
XX
XX AAG19287;
XX
XX 17-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 21031.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX Arabidopsis thaliana.
XX
XX EPI033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-00301439.
XX
XX 25-FEB-1999; 99US-0121825P.
XX
XX 05-MAR-1999; 99US-0123180P.
XX
XX 09-MAR-1999; 99US-0123548P.
XX
XX 23-MAR-1999; 99US-0125788P.
XX
XX 25-MAR-1999; 99US-0126264P.
XX
XX 29-MAR-1999; 99US-0126785P.
XX
XX 01-APR-1999; 99US-0127462P.
XX
XX 06-APR-1999; 99US-0128234P.
XX
XX 08-APR-1999; 99US-0129845P.
XX
XX 16-APR-1999; 99US-0130077P.
XX
XX 19-APR-1999; 99US-0130449P.
XX
XX 21-APR-1999; 99US-0130510P.
XX
XX 23-APR-1999; 99US-0130891P.
XX
XX 28-APR-1999; 99US-0131449P.
XX
XX 30-APR-1999; 99US-0132047P.
XX
XX 04-MAY-1999; 99US-0132484P.
XX
XX 05-MAY-1999; 99US-0132485P.
XX
XX 06-MAY-1999; 99US-0132486P.
XX
XX 07-MAY-1999; 99US-0132487P.
XX
XX 11-MAY-1999; 99US-0132863P.
XX
XX 11-MAY-1999; 99US-0134256P.
XX
XX 14-MAY-1999; 99US-0134218P.
XX
XX 14-MAY-1999; 99US-0134219P.
XX
XX 14-MAY-1999; 99US-0134221P.
XX
XX 14-MAY-1999; 99US-0134370P.
XX
XX 18-MAY-1999; 99US-0134768P.
XX
XX 19-MAY-1999; 99US-0134941P.
XX
XX 20-MAY-1999; 99US-0135124P.
XX
XX 21-MAY-1999; 99US-0135353P.
XX
XX 24-MAY-1999; 99US-0135629P.
XX
XX 25-MAY-1999; 99US-0136021P.
XX
XX 27-MAY-1999; 99US-0136392P.

PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 23-JUN-1999; 99US-0140355P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 02-JUL-1999; 99US-0142154P.
PR 06-JUL-1999; 99US-0142055P.
PR 08-JUL-1999; 99US-0142390P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142927P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.

PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-015138P.
PR 01-SEP-1999; 99US-0152363P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154799P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155559P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 14-OCT-1999; 99US-0159584P.
PR 18-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 22-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 25-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 26-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

PR	02-AUG-1999	99US-0146388P
PR	02-AUG-1999	99US-0146388P
PR	02-AUG-1999	99US-0146389P
PR	03-AUG-1999	99US-0147038P
PR	04-AUG-1999	99US-0147204P
PR	04-AUG-1999	99US-0147302P
PR	05-AUG-1999	99US-0147192P
PR	05-AUG-1999	99US-0147260P
PR	06-AUG-1999	99US-0147703P
PR	06-AUG-1999	99US-0147416P
PR	09-AUG-1999	99US-0147933P
PR	10-AUG-1999	99US-0148171P
PR	11-AUG-1999	99US-0148319P
PR	13-AUG-1999	99US-0148565P
PR	13-AUG-1999	99US-0148684P
PR	16-AUG-1999	99US-0149368P
PR	17-AUG-1999	99US-0149175P
PR	18-AUG-1999	99US-0149426P
PR	20-AUG-1999	99US-0149722P
PR	20-AUG-1999	99US-0149733P
PR	20-AUG-1999	99US-0149929P
PR	22-AUG-1999	99US-0149920P
PR	23-AUG-1999	99US-0149930P
PR	26-AUG-1999	99US-0150566P
PR	26-AUG-1999	99US-0150884P
PR	27-AUG-1999	99US-0151065P
PR	27-AUG-1999	99US-0151080P
PR	30-AUG-1999	99US-0151303P
PR	31-AUG-1999	99US-0151388P
PR	01-SEP-1999	99US-0151930P
PR	07-SEP-1999	99US-0152363P
PR	10-SEP-1999	99US-0153707P
PR	13-SEP-1999	99US-0153785P
PR	15-SEP-1999	99US-0154018P
PR	16-SEP-1999	99US-0154039P
PR	20-SEP-1999	99US-0154779P
PR	22-SEP-1999	99US-0155139P
PR	23-SEP-1999	99US-0155486P
PR	24-SEP-1999	99US-0155569P
PR	28-SEP-1999	99US-0156548P
PR	29-SEP-1999	99US-0156596P
PR	04-OCT-1999	99US-0157117P
PR	05-OCT-1999	99US-0157753P
PR	06-OCT-1999	99US-0157865P
PR	07-OCT-1999	99US-0158029P
PR	08-OCT-1999	99US-0158232P
PR	12-OCT-1999	99US-0158369P
PR	13-OCT-1999	99US-0158963P
PR	13-OCT-1999	99US-0159294P
PR	14-OCT-1999	99US-0159329P
PR	14-OCT-1999	99US-0159330P
PR	14-OCT-1999	99US-0159331P
PR	14-OCT-1999	99US-0159637P
PR	14-OCT-1999	99US-0159638P
PR	18-OCT-1999	99US-0159584P
PR	21-OCT-1999	99US-0160741P
PR	21-OCT-1999	99US-0160767P
PR	22-OCT-1999	99US-0160981P
PR	22-OCT-1999	99US-0160989P
PR	25-OCT-1999	99US-0161404P
PR	25-OCT-1999	99US-0161405P
PR	25-OCT-1999	99US-0161406P
PR	26-OCT-1999	99US-0161359P
PR	26-OCT-1999	99US-0161360P

PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match	25.7%;	Score 319;	DB 3;	Length 439;
Best Local Similarity	36.2%;	Pred. No. 1.5e-26;		
Matches	81;	Conservative	31;	Mismatches 82;
				Indels 30;
				Gaps 5;

Qy 1 LTLALVQGVSDFLILGKSLDILAPGMDLKYARRSLIKHECSTWSGKRIEHLAKGNRF 60
Db 224 LTLVLAHKIGQYTQLSTTVDDAIGFAFDKTAKGLDMH-----RSGPVAEELALBGDAK 278

OY 61HFDIKPLHHAAKNDFSPFGLOHTDVKIMKEEKGQILISSA-----ADIAA 112
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 279SVLENFVEPKYHKDCNFSYAIGLKQVRLALEAKE----IDAKCPSSATHEDRNRADIAA 334

Oy 113 TVQHTMACIIKVKTRHALLFCCKRDLLFNONNAVIVASGCVASINPIIRALIEILTNATQCT 172
 Db 335 SFRVAVLHLEEKCEKRALDWALF---LEPSIKHNVISGCVASNNKYVRLRLNNIVENIKNIK 391

```
QY      173 LTCPPLCTDNGIMAMNGIERLRAGLGILHDIETGRYEKP 216
          ||| | | | | | : | | | | | | | | | | | | | |
Db      392 LVCPPLSLCTDNGVMVAWTGLEHFRVG-----RYDPPPP 425
```

RESULT 13
AAG19285
ID AAG19285 standard: protein: 444 AA

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 21029.

KM Protein identification; signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
KM termination sequence.

OS Arabidopsis thaliana.
XX

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-00301439.

PR	25-FEB-1999;	99US-0121825P.
PR	05-MAR-1999:	99US-0123180P.

PR 23-MAR-1999; 99US-0125788P.
PR 23-MAR-1999; 99US-0125788P.
PR 23-MAR-1999; 99US-0125788P.

PR	29-MAR-1999;	99US-0126785P.
PR	01-APR-1999:	99US-0127462P.

PR 08-APR-1999; 99US-0128714P.
PR 08-APR-1999; 99US-0128234F.

PR	19-APR-1999;	99US-0130077P.
PR	21-APR-1999;	99US-0130449P.

PR 23-APR-1999; 99US-0130891P

PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.

PR	05-MAY-1999;	99US-0132485P
EN	04-MAY-1999;	99US-0132485P
PR	05-MAY-1999;	99US-0132485P
EN	04-MAY-1999;	99US-0132485P

PR	06-MAY-1999;	99US-0132487P.
PR	07-MAY-1999;	99US-0132863P.

PR 14-MAY-1999; 99US-0134218P

Query Match	25.7%; Score 319; DB 3; Length 439;
Best Local Similarity	36.2%; Pred. No.1.5e-26;
Matches	81; Conservative 31; Mismatches 82; Indels 30; Gaps 5
Qy 1	LLALVGVSPDFLLIGKSLIDAPGDMLEKVARRLSLIKHPECSYMSGGKALIEHLAKQGNRF 60
Dd 224	LLVLAHKLGGYTGQGTGTVDDALIGAFPKXTAKWLGIDMH-----RSGCPAVEBLALEGDAK 278
Qy 61	HFIDIKPRLHAAKNCDFSEFTGLQHTVDKIIMKKEKEEGIEKQILLSA-----ADTAA 112
Dd 279	SVKENVEMPKYHKDCNCFSYAGLKTQVRLAIEAKE----IDACCPVSSATNEDRRNRADIAA 334
Qy 113	TVQHTMACHLVKRTHRAILFEKQKRDLLPQNNAVLVASGVSANFYIRPALIELTNATQCT 172
Dd 335	SFGQVAVVLHLEKEKERALDMLB--LEPSIKHWVISGGVANSKTVRLPLNNIVENKNIK 391
Qy 173	LLCEPPLCTDNGIMIMAMNGIERLRAGIGILHDIEGIRYEPKCP 216
Dd 392	LVCPPSLCTDNGYVAVMTGIEHRRVG-----RYDEPPP 425
RESULT 13	
AAG19285	
ID	AAG19285 standard; protein; 444 AA.
AC	
XX	AAG19285;
DT	17-OCT-2000 (first entry)
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 21029.
KW	Protein identification; signal transduction pathway; metabolic pathway;
KM	hybridisation assay; genetic mapping; gene expression control; promoter;
XX	termination sequence.
OS	Arabidopsis thaliana.
XX	
PN	EP1033405-A2.
PD	
XX	06-SEP-2000.
PF	
XX	25-FEB-2000; 2000EP-00301439.
PR	99US-0121825P.
PR	99US-012180P.
PR	99US-0123548P.
PR	99US-0125788P.
PR	99US-0126264P.
PR	99US-0126785P.
PR	99US-0127462P.
PR	99US-0128234P.
PR	99US-0128714P.
PR	99US-0129845P.
PR	99US-0130077P.
PR	99US-0130449P.
PR	99US-0130519P.
PR	99US-0130891P.
PR	99US-0132484P.
PR	99US-0132485P.
PR	99US-0132487P.
PR	99US-0132486P.
PR	99US-0132863P.
PR	99US-0134256P.
PR	99US-0134218P.

PR 14-MAY-1999;	99US-0134219P.	PR 26-JUL-1999;	99US-0145276P.
PR 14-MAY-1999;	99US-0134222P.	PR 27-JUL-1999;	99US-0145913P.
PR 14-MAY-1999;	99US-0134370P.	PR 27-JUL-1999;	99US-0145918P.
PR 18-MAY-1999;	99US-0134768P.	PR 27-JUL-1999;	99US-0145919P.
PR 19-MAY-1999;	99US-0134941P.	PR 28-JUL-1999;	99US-0145951P.
PR 20-MAY-1999;	99US-0135124P.	PR 02-AUG-1999;	99US-0146386P.
PR 21-MAY-1999;	99US-0135353P.	PR 02-AUG-1999;	99US-0146388P.
PR 24-MAY-1999;	99US-0135629P.	PR 02-AUG-1999;	99US-0146389P.
PR 25-MAY-1999;	99US-0136021P.	PR 03-AUG-1999;	99US-0147038P.
PR 27-MAY-1999;	99US-0136392P.	PR 04-AUG-1999;	99US-0147204P.
PR 28-MAY-1999;	99US-0136782P.	PR 04-AUG-1999;	99US-0147302P.
PR 01-JUN-1999;	99US-0137222P.	PR 05-AUG-1999;	99US-0147192P.
PR 03-JUN-1999;	99US-0137528P.	PR 05-AUG-1999;	99US-0147260P.
PR 04-JUN-1999;	99US-0137502P.	PR 06-AUG-1999;	99US-0147303P.
PR 07-JUN-1999;	99US-0137724P.	PR 06-AUG-1999;	99US-0147416P.
PR 08-JUN-1999;	99US-0138094P.	PR 09-AUG-1999;	99US-0147933P.
PR 10-JUN-1999;	99US-0138540P.	PR 09-AUG-1999;	99US-0147935P.
PR 10-JUN-1999;	99US-0138647P.	PR 10-AUG-1999;	99US-0148171P.
PR 14-JUN-1999;	99US-0139112P.	PR 11-AUG-1999;	99US-0148319P.
PR 16-JUN-1999;	99US-0139452P.	PR 12-AUG-1999;	99US-0148341P.
PR 16-JUN-1999;	99US-0139453P.	PR 13-AUG-1999;	99US-0148565P.
PR 17-JUN-1999;	99US-0139492P.	PR 13-AUG-1999;	99US-0148684P.
PR 18-JUN-1999;	99US-0139454P.	PR 16-AUG-1999;	99US-0149368P.
PR 18-JUN-1999;	99US-0139456P.	PR 17-AUG-1999;	99US-0149175P.
PR 18-JUN-1999;	99US-0139457P.	PR 18-AUG-1999;	99US-0149426P.
PR 18-JUN-1999;	99US-0139458P.	PR 20-AUG-1999;	99US-0149722P.
PR 18-JUN-1999;	99US-0139459P.	PR 20-AUG-1999;	99US-0149723P.
PR 18-JUN-1999;	99US-0139460P.	PR 20-AUG-1999;	99US-0149929P.
PR 18-JUN-1999;	99US-0139461P.	PR 23-AUG-1999;	99US-0149902P.
PR 18-JUN-1999;	99US-0139462P.	PR 23-AUG-1999;	99US-0149930P.
PR 18-JUN-1999;	99US-0139463P.	PR 25-AUG-1999;	99US-0150566P.
PR 18-JUN-1999;	99US-0139750P.	PR 26-AUG-1999;	99US-0150884P.
PR 18-JUN-1999;	99US-0139763P.	PR 27-AUG-1999;	99US-0151065P.
PR 21-JUN-1999;	99US-0139817P.	PR 27-AUG-1999;	99US-0151066P.
PR 22-JUN-1999;	99US-0139899P.	PR 30-AUG-1999;	99US-0151080P.
PR 23-JUN-1999;	99US-0140353P.	PR 31-AUG-1999;	99US-0151303P.
PR 23-JUN-1999;	99US-0140354P.	PR 01-SEP-1999;	99US-0151338P.
PR 24-JUN-1999;	99US-0140695P.	PR 07-SEP-1999;	99US-0151309P.
PR 28-JUN-1999;	99US-0140823P.	PR 10-SEP-1999;	99US-0152363P.
PR 29-JUN-1999;	99US-0140991P.	PR 13-SEP-1999;	99US-0153070P.
PR 30-JUN-1999;	99US-0141287P.	PR 15-SEP-1999;	99US-0153758P.
PR 01-JUL-1999;	99US-0141842P.	PR 16-SEP-1999;	99US-0154018P.
PR 01-JUL-1999;	99US-0142154P.	PR 20-SEP-1999;	99US-0154039P.
PR 02-JUL-1999;	99US-0142055P.	PR 22-SEP-1999;	99US-0154479P.
PR 06-JUL-1999;	99US-0142399P.	PR 23-SEP-1999;	99US-0155139P.
PR 08-JUL-1999;	99US-0142803P.	PR 24-SEP-1999;	99US-0155486P.
PR 09-JUL-1999;	99US-0142920P.	PR 28-SEP-1999;	99US-0155659P.
PR 12-JUL-1999;	99US-0143524P.	PR 29-SEP-1999;	99US-0156458P.
PR 13-JUL-1999;	99US-0143624P.	PR 04-OCT-1999;	99US-0157117P.
PR 14-JUL-1999;	99US-0144005P.	PR 05-OCT-1999;	99US-0157753P.
PR 16-JUL-1999;	99US-0144085P.	PR 06-OCT-1999;	99US-0157865P.
PR 16-JUL-1999;	99US-0144086P.	PR 07-OCT-1999;	99US-0158029P.
PR 19-JUL-1999;	99US-0144325P.	PR 08-OCT-1999;	99US-0158232P.
PR 19-JUL-1999;	99US-0144331P.	PR 12-OCT-1999;	99US-0158369P.
PR 19-JUL-1999;	99US-0144332P.	PR 13-OCT-1999;	99US-0158923P.
PR 19-JUL-1999;	99US-0144333P.	PR 13-OCT-1999;	99US-0159294P.
PR 19-JUL-1999;	99US-0144334P.	PR 13-OCT-1999;	99US-0159295P.
PR 19-JUL-1999;	99US-0144335P.	PR 14-OCT-1999;	99US-0159329P.
PR 20-JUL-1999;	99US-0144336P.	PR 14-OCT-1999;	99US-0159330P.
PR 20-JUL-1999;	99US-0144633P.	PR 14-OCT-1999;	99US-0159331P.
PR 20-JUL-1999;	99US-0144684P.	PR 14-OCT-1999;	99US-0159637P.
PR 21-JUL-1999;	99US-0144814P.	PR 14-OCT-1999;	99US-0159638P.
PR 21-JUL-1999;	99US-0145086P.	PR 18-OCT-1999;	99US-0159584P.
PR 21-JUL-1999;	99US-0145088P.	PR 21-OCT-1999;	99US-0160741P.
PR 22-JUL-1999;	99US-0145085P.	PR 21-OCT-1999;	99US-0160767P.
PR 22-JUL-1999;	99US-0145087P.	PR 21-OCT-1999;	99US-0160768P.
PR 22-JUL-1999;	99US-0145089P.	PR 21-OCT-1999;	99US-0160770P.
PR 22-JUL-1999;	99US-0145152P.	PR 21-OCT-1999;	99US-0160814P.
PR 23-JUL-1999;	99US-0145145P.	PR 21-OCT-1999;	99US-0160815P.
PR 23-JUL-1999;	99US-0145218P.	PR 22-OCT-1999;	99US-0160980P.
PR 23-JUL-1999;	99US-0145224P.	PR 22-OCT-1999;	99US-0160981P.
		PR 22-OCT-1999;	99US-0160989P.

PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161921P.
PR 28-OCT-1999; 99US-0161931P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 25.7%; Score 319; DB 3; Length 444;
Best Local Similarity 36.2%; Pred. No. 1.5e-26;
Matches 81; Conservative 31; Mismatches 82; Indels 30; Gaps 5;

QY 1 LIALVGVSPDLLGKSLDAPGMDLVARSLIKPECTSGGKAIETLAKQGNRF 60
DB 229 LVLIAHKIGQYTOGLTVDDAIGFAFDYAKWGLDMH-----RSGGPAVEELALEGDAK 283
QY 61 HFDIKPLHAKNCDFSTGLOHVNVDKLIIMKKEEGIEKGQILSSA-----ADIAA 112
DB 284 SVKKNVPMKXHKDCNFSYAGLKTQVRLAIEAKG---IDAKCPVSATMEDRRRADIAA 339
QY 113 TVQHTMACHLVKTTHRAILFCQKQDLPQNNNAVVASGVASNFYIRALIELTNATQCT 172
DB 340 SFQVAVVHLHEKCKRAIDWALE--LEPSIKHWISGVASNAKCVRLANNIVENKOLK 396
QY 173 LLCPPRLCTDNGIMIANWGIERLRAGIGILHDIGIRPEKCP 216
DB 397 LVCPPEPRLCTDNGVAVMTGLEHFRVG-----RYDPPPP 430

RESULT 14
ABU22934

ID ABU22934 standard; protein; 382 AA.

XX AC ABU22934;

XX DT 19-JUN-2003 (first entry)

XX DE Protein encoded by Prokaryotic essential gene #8461.

XX KM Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX OS Bordetella pertussis.

XX PN MO200277183-A2.

PD 03-OCT-2002.

XX PF 21-MAR-2002; 2002WO-US009107.

XX PR 21-MAR-2001; 2001US-00815242.

XX PR 06-SEP-2001; 2001US-00948993.

XX PR 25-OCT-2001; 2001US-0342923P.

XX PR 08-FEB-2002; 2002US-00072851.

XX PR 06-MAR-2002; 2002US-0362699P.

XX PA (ELIT-) ELITRA PHARM INC.

XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zykkind JW;

XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX DR WPI; 2003-029926/02.

XX DR N-PSDB; ACA26804.

XX PT New antisense nucleic acids, useful for identifying proteins or screening
XX PT for homologous nucleic acids required for cellular proliferation to
XX PT isolate candidate molecules for rational drug discovery programs.
XX PS Claim 25; SEQ ID NO 50858; 1766pp; English.
XX CC The invention relates to an isolated nucleic acid comprising any one of

CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 382 AA;

Query Match 20.7%; Score 257; DB 6; Length 382;
Best Local Similarity 36.1%; Pred. No. 1.2e-19;
Matches 74; Conservative 33; Mismatches 78; Indels 20; Gaps 7;

QY 2 LALVGVSPDLLGKSLDAPGMDLVARSLIKPECTSGGKAIETLAKQGNRF 61

DB 178 LMLVDGVGRYELLETLDNAGAFPKSAKMLG-CYP-----GCPALRLALEQGDASR 230

QY 62 FDIKPLHAKNCDFSTGLO-HVTDKLIIMKKEEGIEKGQILSSAADIAATVQHTMAC 120

DB 231 YDLPRMLHSGDLDFSGKTRVLRV--KATRDGGEIGR--QBRADIAATQAIVE 286

QY 121 HLVKRTTHRAILFCQKQDLPQNNNAVVASGVASNFYIRALIELTNATQCTLLCPPRL 180

DB 287 VLAAKAIKRAL--KQTGL-----RLVNAVGVAGVANNLRHLRALKPLAAEAYFPPLSL 338

QY 181 CTDNGIMIANWGIERLRAGIGILHD 205

DB 339 CTDNGAMIAFPAAEKRYKGLADLRE 363

RESULT 15
ABU35232
ID ABU35232 standard; protein; 348 AA.

XX AC ABU35232;

XX DT 19-JUN-2003 (first entry)

XX DE Protein encoded by Prokaryotic essential gene #20759.

XX KM Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX OS Moraxella catarrhalis.

XX PN MO200277183-A2.

XX PD 03-OCT-2002.

XX PF 21-MAR-2002; 2002WO-US009107.

DB 306 CTDNGMIAVAGFCRLSCG 324
 Search completed: November 10, 2005, 16:48:03
 Job time : 65.4261 secs

XX 21-MAR-2001; 2001US-00815242.
 PR 06-SEP-2001; 2001US-00948993.
 PR 25-OCT-2001; 2001US-0342923P.
 PR 08-FEB-2002; 2002US-00072851.
 PR 06-MAR-2002; 2002US-0362699P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zykkind JW;
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Foreysch RA, Xu HH;
 XX
 DR WPI: 2003-029926/02.
 DR N-PSDB; ACA39102.
 XX
 PT New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.
 PS
 PS Claim 25; SEQ ID NO 63156; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pat_sequences

XX Sequence 348 AA;

Query Match 20.0%; Score 248.5; DB 6; Length 348;
 Best Local Similarity 33.2%; Pred. No. 9.4e-19;
 Matches 66; Conservative 33; Mismatches 79; Indels 21; Gaps 5;

QY 1 LLAALVGVSDPLLLKSLIDIAAGDMLDKVARRLSLIKAPBCSTWSGRAIEHLAKQGRF 60
 DB 147 MLVRADGVGVYQIIIGESIDDAVGECPDKTAKMLTKLPP-----GPNIEKLAQNGNPH 199
 QY 61 HPDIKPLHLAANKCDFSTGTGLOHTVDKLIKKEKEEGIEKGQIISAAADIAATVOHTMAC 120
 DB 200 AYEELPRPMOH-KGLDFSPSGKTAIHNLIKDTPNQSPP-----ATRADIAASFVAVD 253
 QY 121 HLVRGTRAILFCQKQDRLPQNNAVLVASGVASFYIRBALBITNATQCTLLCPPRL 180
 DB 254 TLVKKCTVALQMTGIRNQ-----LVVAGVSANQPLRKLITLIRQIDASVITAPTEL 305
 QY 181 CTDNGMIAVAGFCRLRAG 199
 ||||| ||| | | |

C;Accession: AB2902
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.
erage, G.; Gillet, M.; Grant, C.; Guenther, D.; Kuyavlin, T.; Levy, R.; Li, M.; McClie
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Kreepan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer *Agrobacterium tumefaciens* C58.
A;Reference number: AB2577; MUID:21608550; PMID:11743193
A;Accession: AB2902
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-365 <KUR>
A;Cross-references: GB:AE006688; PIDN:AA143632.1; PID:g17741154; GSPDB:GN00186
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: gcp
A;Map position: circular chromosome
C;Superfamily: O-sialoglycoprotein endopeptidase

Query Match 19.7%; Score 244; DB 2; Length 365;
Best Local Similarity 31.9%; Pred. No. 1.1e-14;
Matches 73; Conservative 34; Mismatches 92; Indels 30; Gaps 7;

Qy 2 LALVGVSDPFLILGKSLDIPGMDLVKVARSLIKHPGCTSMGSKATEHLAKOGRFH 61
Db LVLVRGVEYERWGTITDIALGEAFDPTAKLGL-PPY-----GGPVENMAAKGDPR 201
149 LVLVRGVEYERWGTITDIALGEAFDPTAKLGL-PPY-----GGPVENMAAKGDPR 201

Qy 62 FDIKPELHNAKCDPFSFTGL-----QHTVDKIIMKEKEGIEKGQILSSADIATVQHT 117
Db FPLRPFWGGEARLDPSFGLKTAVAQAATAIAPLSEOD-----IADICASFQRA 250
202 FPLRPFWGGEARLDPSFGLKTAVAQAATAIAPLSEOD-----IADICASFQRA 250

Qy 118 MACHLVKTRTHAILFCCKRDLLPQNN--VLVAGGVASNFYIRALLETITNATQCTLLC 175
Db VSRILKORIGRGLARFKE--PPHNGEPALVAVAGVANOSIRQLQLCPTHGRFVA 308
251 VSRILKORIGRGLARFKE--PPHNGEPALVAVAGVANOSIRQLQLCPTHGRFVA 308

Qy 176 PPRRLCTDNGIMIANWGIERLAGLILHDEIGIRYEPKCPGVNISKE 224
Db PPHRLCTDNAMIAMWAGIERMAEG-----ROADALVAVAPRSRWPDLGSAE 353
309 PPHRLCTDNAMIAMWAGIERMAEG-----ROADALVAVAPRSRWPDLGSAE 353

RESULT 3
D97677
probable o-sialoglycoprotein endopeptidase (glycoproteinase) (imported) - *Agrobacterium*
C;Species: *Agrobacterium tumefaciens*
C;Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 18-Nov-2002
C;Accession: D97677
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughly, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent *Agrobacterium tum*
A;Reference number: A97359; MUID:21608551; PMID:11743194
A;Accession: D97677
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-366 <KUR>
A;Cross-references: GB:AE007869; PIDN:AAK8373.1; PID:g15157858; GSPDB:GN00169
C;Genetics:
A;Gene: AGR C 4806
A;Map position: circular chromosome
C;Superfamily: O-sialoglycoprotein endopeptidase

Query Match 19.7%; Score 244; DB 2; Length 366;
Best Local Similarity 31.9%; Pred. No. 1.1e-14;
Matches 73; Conservative 34; Mismatches 92; Indels 30; Gaps 7;

Qy 2 LALVGVSDPFLILGKSLDIPGMDLVKVARSLIKHPGCTSMGSKATEHLAKOGRFH 61
Db LVLVRGVEYERWGTITDIALGEAFDPTAKLGL-PPY-----GGPVENMAAKGDPR 202
150 LVLVRGVEYERWGTITDIALGEAFDPTAKLGL-PPY-----GGPVENMAAKGDPR 202

Qy 62 FDIKPELHNAKCDPFSFTGL-----QHTVDKIIMKEKEGIEKGQILSSADIATVQHT 117
Db FPLRPFWGGEARLDPSFGLKTAVAQAATAIAPLSEOD-----IADICASFQRA 251
203 FPLRPFWGGEARLDPSFGLKTAVAQAATAIAPLSEOD-----IADICASFQRA 251

Qy 118 MACHLVKTRTHAILFCCKRDLLPQNN--VLVAGGVASNFYIRALLETITNATQCTLLC 175
Db VSRILKORIGRGLARFKE--PPHNGEPALVAVAGVANOSIRQLQLCPTHGRFVA 309
251 VSRILKORIGRGLARFKE--PPHNGEPALVAVAGVANOSIRQLQLCPTHGRFVA 309

Qy 176 PPRRLCTDNGIMIANWGIERLAGLILHDEIGIRYEPKCPGVNISKE 224
Db PPHRLCTDNAMIAMWAGIERMAEG-----ROADALVAVAPRSRWPDLGSAE 354
310 PPHRLCTDNAMIAMWAGIERMAEG-----ROADALVAVAPRSRWPDLGSAE 354

RESULT 4
AB3274
O-sialoglycoprotein endopeptidase (BC 3.4.24.57) (imported) - *Brucella melitensis* (strai
C;Species: *Brucella melitensis*
C;Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 15-Feb-2002
C;Accession: AB3274
R;DelVecchio, V.G.; Kapatal, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,
.; Mazur, M.; Goldsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A;Title: The genome sequence of the facultative intracellular pathogen *Brucella melitensis*
A;Reference number: AD3252; PMID:11756688
A;Accession: AB3274
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-359 <KUR>
A;Cross-references: GB:AE008917; PIDN:AA151357.1; PID:g17982056; GSPDB:GN00190
A;Experimental source: strain 16M
C;Genetics:
A;Gene: BME10175
A;Map position: I
C;Superfamily: O-sialoglycoprotein endopeptidase
C;Keywords: hydrolase; metalloproteinase

Query Match 19.5%; Score 241.5; DB 2; Length 359;
Best Local Similarity 34.5%; Pred. No. 1.9e-14;
Matches 69; Conservative 27; Mismatches 77; Indels 27; Gaps 6;

Qy 2 LALVGVSDPFLILGKSLDIPGMDLVKVARSLIKHPGCTSMGSKATEHLAKOGRFH 61
Db LVLVRGVEYERWGTITDIALGEAFDPTAKLGL-PPY-----GGPVENMAAKGDPR 197
145 LVLVRGVEYERWGTITDIALGEAFDPTAKLGL-PPY-----GGPVENMAAKGDPR 197

Qy 62 FDIKPELHNAKCDPFSFTGL-----QHTVDKIIMKEKEGIEKGQILSSADIATVQHT 116
Db FPLRPFWGGEARLDPSFGLKTAVAQAATAIAPLSEOD-----IADICASFQRA 245
198 FPLRPFWGGEARLDPSFGLKTAVAQAATAIAPLSEOD-----IADICASFQRA 245

Qy 117 MACHLVKTRTHAILFCCKRDLLPQ--NNVLVAGGVASNFYIRALLETITNATQCTLLC 175
Db AVADTSLDSVGRSLERFKE--PPDCATPSLVVAGVANAKTIRALLETITNATQCTLLC 303
246 AVADTSLDSVGRSLERFKE--PPDCATPSLVVAGVANAKTIRALLETITNATQCTLLC 303

Qy 176 PPRRLCTDNGIMIANWGIER 195
Db PPLNCTDNAMIAMWAGIER 323
304 PPLNCTDNAMIAMWAGIER 323

RESULT 5
E71711
probable o-sialoglycoprotein endopeptidase (gcp) RP037 - *Rickettsia prowazekii*
C;Species: *Rickettsia prowazekii*
C;Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 09-Jul-2004
C;Accession: E71711
R;Andersson, S.G.B.; Zomrodipour, A.; Andersson, J.O.; Sichertz-Ponten, T.; Alsmark, U
Nature 396, 133-140, 1998
A;Title: The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria.
A;Reference number: A71630; MUID:99039499; PMID:9823893
A;Accession: E71711
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-387 <AND>
A;Cross-references: UNIPROT:Q9ZEA8; GB:AJ235270; GB:AJ235269; NID:g3860572; PIDN:CAA1450
A;Experimental source: strain Madrid E
C;Genetics:
A;Gene: gcp; RP037
C;Superfamily: O-sialoglycoprotein endopeptidase

Query Match 19.2%; Score 238; DB 2; Length 387;
Best Local Similarity 29.7%; Pred. No. 4.4e-14;
Matches 70; Conservative 32; Mismatches 74; Indels 60; Gaps 6;

5 VGVSDPFLLGKSLDIAPGMDLKVARRLSLIKHEPCSTMSGKALIEHLAKGNRHPDI 64
145 VLGIKTKIILGTTIDDAVGEFDPKVAKNLNL-----SFGPEIEKRAKLGPHKXKF 197
65 KPPLHAAKNCDSPFTGLQHTVDKIIIMK-KEKEEGIEKGQILSSAADIAATVQHTMACHLV 123
198 KPPIINSNCNMSFSGLTAVRTLIIMLKAVNDV-----INDIASPFTIGALIS 249

124 KRTIRALIFCKQ-----RD-----LLPON---- 142
250 SKQODAIRLYKQIINDYEDINHPKLNLSKFRKDEPFNMKPLECTIRPKYRIHIQNSYRS 309

143 ---NAVVASGVSANFYIRALBITLMACTCTLCPPRLCTDNGMIAMNGIER 195
310 NLNDITVIAGVANKYLOEILSDCTRPYGRILIAPEMHLCTDNAAAMIAVAGLER 365

RESULT 6
F87257
peptidase M22 family protein [imported] - Caulobacter crescentus
C:/Species: Caulobacter crescentus
C:/Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C:/Accession: F87257

R;Nieman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.B.; Laub, M.T.; Deboy, R.T.; Dodson, R.U.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kjoltor, n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M. Proc. Natl. Acad. Sci. U.S.A. 98, 4133-4141, 2001

A:/Title: Complete Genome Sequence of Caulobacter crescentus.
A:/Reference number: A87249; WUID:21173698; PMID:11259647
A:/Accession: F87257
A:/Status: preliminary
A:/Molecule type: DNA
A:/Residues: 1-367 <STO>
A:/Cross-references: GB:AE005673; NID:G13421168; PIDN:AAK22058.1; GSPDB:GN00148
C:/Genetics:
A:/Gene: CC0071
C:/Superfamily: O-sialoglycoprotein endopeptidase

Query Match 19.2%; Score 237.5; DB 2; Length 367;
Best Local Similarity 34.3%; Pred. No. 4.6e-14;
Matches 73; Conservative 29; Mismatches 90; Indels 21; Gaps 6;

5 VGVSDPFLLGKSLDIAPGMDLKVARRLSLIKHEPCSTMSGKALIEHLAKGNRHPDI 64
156 VSGVGACRRLGTTIDDAAGEAFDKIAKSLGL-PYP-----GCPALREKLVAGDPTRYAL 208

65 KPPLHAAKNCDSPFTGLQHTVDKIIIMKKEKEEGIEKGQILSSAADIAATVQHTMACHLV 124
209 PRALLGRKDCDFSGKTAARJAETLTTDD-----ARBLAAGVQAAIARQISE 259

125 RTRRALIFCKQBDLPQNNNAVIVASGVANFYIRALBITLMACTCTLCPPRLCTDN 184
260 RVRBAKLYK--DSHDEDLAFVAVAGVAAVAALADCEKNGFSFAAPPLAYCTDN 317

185 GIMIAMNGIERLRAGLILHDIGIEGRYEPKPL 217
318 AAMIALAGARL--ALGIFDDLDAIA-RPRWPL 347

RESULT 7
H83572
O-sialoglycoprotein endopeptidase PA0580 [imported] - Pseudomonas aeruginosa (strain PAO
C:/Species: Pseudomonas aeruginosa
C:/Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:/Accession: H83572

R;Stover, C.K.; Pham, X.Q.; Brwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Bz adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Ladhig, K.; Lim, .; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:/Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:/Reference number: A82950; WUID:20437337; PMID:10984043
A:/Accession: H83572
A:/Status: preliminary
A:/Molecule type: DNA
A:/Residues: 1-341 <STO>
A:/Cross-references: GB:AE004494; GB:AE004091; NID:g9946446; PIDN:AA03969.1; GSPDB:GN001
C:/Genetics:
A:/Gene: gcp
C:/Superfamily: O-sialoglycoprotein endopeptidase

Query Match 18.6%; Score 231; DB 2; Length 341;
Best Local Similarity 33.8%; Pred. No. 1.6e-13;
Matches 73; Conservative 34; Mismatches 87; Indels 22; Gaps 8;

2 LALVGVSDPFLLGKSLDIAPGMDLKVARRLSLIKHEPCSTMSGKALIEHLAKGNRHPDI 61
142 LVRVIGRYQLGSDVDADAGEAFDKIAKSLGL-GYP-----GPEIARLIERGTGPR 194

62 FDIKPPLHAAKNCDSPFTGLQHTVDKIIIMKKEKEEGIEKGQILSSAADIAATVQHTMACH 121
195 FVFPFPMTPDRPGLDPSFSLKTFITLN-TWQRCVAGDDBEQ---TRCDIALAFQTAVERT 250

122 LVKTRRALIFCKQBDLPQNNNAVIVASGVANFYIRALBITLMACTCTLCPPRLCT 181
251 LIIKRRAL--KQGL--KN---LVYAGVSANQALRSGLEKMGEMKGVYARPRC 302

182 TDNGMIAMNGIERLRAGLILHDIGIEGRYEPKPL 217
303 TDNGAMIAVAGCORLLAG---QHDGPALISVOPRPMW 335

RESULT 8
E97707
O-sialoglycoprotein endopeptidase (EC 3.4.24.57) [imported] - Rickettsia conorii (strain
C:/Species: Rickettsia conorii
C:/Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 22-Oct-2001
C:/Accession: E97707

R;Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; Ro science 293, 2093-2098, 2001

A:/Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
A:/Reference number: A97700; WUID:21442074; PMID:11557893
A:/Accession: E97707
A:/Status: preliminary
A:/Molecule type: DNA
A:/Residues: 1-344 <KUR>
A:/Cross-references: GB:AE006914; PIDN:AAL02599.1; PID:g15619097; GSPDB:GN00173
C:/Genetics:
A:/Gene: gcp
C:/Superfamily: O-sialoglycoprotein endopeptidase
C:/Keywords: hydrolase; metalloproteinase

Query Match 18.6%; Score 230.5; DB 2; Length 344;
Best Local Similarity 33.2%; Pred. No. 1.9e-13;
Matches 64; Conservative 34; Mismatches 76; Indels 19; Gaps 5;

5 VGVSDPFLLGKSLDIAPGMDLKVARRLSLIKHEPCSTMSGKALIEHLAKGNRHPDI 64
145 VLGIKTKIILGTTIDDAVGEAFDKVAKNLNL-----AFPGPEIEKRAKLGDPHXKF 197

65 KPPLHAAKNCDSPFTGLQHTVDKIIIMK-KEKEEGIEKGQILSSAADIAATVQHTMACHLV 123
198 KPPIINSNCNMSFSGLTAVRTLIIMLKAVNDV-----INDIASPFTIGALIS 249

124 KRTIRALIFCKQ-BLLQNNNAVIVASGVANFYIRALBITLMACTCTLCPPRLCT 182
250 SKQODAIRLYKQIINDYEDINHPKLNLSKFRKDEPFNMKPLECTIRPKYRIHIQNSYRS 307

183 DNGIMIAMNGIER 195
308 DNAAAMIAVAGLER 320

A:Cross-references: UNIPROT:066986; GB:AE000708; NID:g2983356; PIDN:AAC06951.1; PID:g298
A:Experimental source: strain VFS
C:Genetics:
A:Gene: gcp
C:Superfamily: O-sialoglycoprotein endopeptidase

Query Match 17.0%; Score 210.5; DB 2; Length 335;
Best Local Similarity 32.0%; Pred. No. 1.2e-11;
Matches 70; Conservative 41; Mismatches 73; Indels 35; Gaps 10;

QY 2 LALVGVSDPFLILGSLDIPAGMDLKVARRSLIKHPECSTMSGGKAIENHLAKOGRPH 61
DB 141 LYLVRDFRYPDLGGTLDVAEAYDKVAKMLGL-GYF-----GGPIIDRLAKESKKU- 192

QY 62 FDIKPEPLHAKNCDPSFTGLQHVTDKIIMKKEKEGIEKGQILSSAADIATVQHTMAC 121
DB 193 YLPKPEPLMEEGNLNFSGLK---TALNLLKKEKNVR-----EDLAYSFOETVARI 242

QY 122 LVKRTHRALPCKQDRLPQNNNAVVASGVASNFYIRALBITLNATQCTLLCPPRL 178
DB 243 LLEKS---LWAMKKTGIRK---LVVVGVSANSRLR---EVFKASQEGFELYIPHP 291

QY 179 RLCTDNGIMIAMNGIERLAGLILHDIEGIRYEPKCP 217
DB 292 SLSTDNMLMIAVAGMERKGVADLVNP--QPNITPL 327

RESULT 13
A:00079
probable glycoproteinase gcp [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C>Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Nov-2001
C:Accession: A10079
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titchall, R.W.; Holden, M.T.G.; Prentice, M.B.
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall,
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11563360
A:Accession: A10079
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-337 <KUR>
A:Cross-references: GB:AL590842; PIDN:CAC89500.1; PID:g15978736; GSPDB:GN00175
C:Genetics:
A:Gene: gcp
C:Superfamily: O-sialoglycoprotein endopeptidase

Query Match 16.9%; Score 210; DB 2; Length 337;
Best Local Similarity 30.0%; Pred. No. 1.4e-11;
Matches 65; Conservative 33; Mismatches 91; Indels 28; Gaps 6;

QY 2 LALVGVSDPFLILGSLDIPAGMDLKVARRSLIKHPECSTMSGGKAIENHLAKOGRPH 61
DB 142 LISVTGIGYEYLLIGESVDADAGEAFDXTAKLLGL-DYP-----GGPMLSRMAOQTVGR 194

QY 62 FDIKPEPLHAKNCDPSFTGLQHVTDKIIMKKEKEGIEKGQILSSAADIATVQHTMAC 121
DB 195 FTFPRPMTDRPGLDPSFGSLKTFPAANTIRANGDD-----QTRADIARAFEDAVDT 246

QY 122 LVKRTHRALPCKQDRLPQNNNAVVASGVASNFYIRALBITLNATQCTLLCPPRL 180
DB 247 LAISKKA-----LDQGFKRLVITAGVSAQOTLRLKLAADMOKRGSGVFPARBP 297

QY 181 CTDNGIMIAMNGIERLAGLILHDIEGIRYEPKCP 217
DB 298 CTDNGAMIAVAGMRLSN---LNSLSVSRPRPPL 331

RESULT 14
F75029
O-sialoglycoprotein endopeptidase (gcp) PAB1159 - Pyrococcus abyssi (strain Orsay)
C:Species: Pyrococcus abyssi

C>Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jul-2004
C:Accession: F75029
R:anonymous, Genoscope
A:Submitted to the EMBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome str
A:Reference number: A75001
A:Accession: F75029
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-324 <KAM>
A:Cross-references: UNIPROT:Q9UXT7; GB:AJ248288; GB:AL096836; NID:g5458960; PIDN:CAB506
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: gcp; PAB1159
C:Superfamily: O-sialoglycoprotein endopeptidase

Query Match 16.9%; Score 209; DB 2; Length 324;
Best Local Similarity 32.5%; Pred. No. 1.6e-11;
Matches 65; Conservative 34; Mismatches 69; Indels 32; Gaps 8;

QY 1 LALVGVSDPFLILGSLDIPAGMDLKVARRSLIKHPECSTMSGGKAIENHLAKOGRPH 60
DB 135 VLALBEG--RNVFGTLDIGNALDVPAREGL-----GPGGPKVEKLAKEGKY 185

QY 61 HDIKPEPLHAKNCDPSFTGLQHVTDKIIMKKEKEGIEKGQILSSAADIATVQHTMAC 120
DB 186 ---IELP-YAVKGMDSFGL--LTAIRKRYSGKRV-----DLAYSFOETAPA 230

QY 121 LVKRTHRALPCKQDRLPQNNNAVVASGVASNFYIRALBITLNATQCTLLCPPRL 180
DB 231 ALVEYTERAVANTE-----KDEVLV--GGVANNRLREMLRIMTEDRGKFEVPPYDL 282

QY 181 CTDNGIMIAMNGIERLAGL 200
DB 283 CRDNGAMIAVITGLRMYKAGI 302

RESULT 15
C91122
probable O-sialoglycoprotein endopeptidase [imported] - Escherichia coli (strain O157:H7)
C:Species: Escherichia coli
C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C:Accession: C91122
R:Havashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gaawara, N.; Yasunaga, T.; Kihara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: C91122
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-337 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA037370.1; PID:g13363420; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain R1MD 0509952
C:Genetics:
A:Gene: BCE3947
C:Superfamily: O-sialoglycoprotein endopeptidase

Query Match 16.9%; Score 209; DB 2; Length 337;
Best Local Similarity 30.6%; Pred. No. 1.7e-11;
Matches 66; Conservative 34; Mismatches 90; Indels 26; Gaps 8;

QY 2 LALVGVSDPFLILGSLDIPAGMDLKVARRSLIKHPECSTMSGGKAIENHLAKOGRPH 61
DB 142 LISVTGIGYEYLLIGESVDADAGEAFDXTAKLLGL-DYP-----GGPLSKMAAOCTAGR 194

QY 62 FDIKPEPLHAKNCDPSFTGLQHVTDKIIMKKEKEGIEKGQILSSAADIATVQHTMAC 121
DB 195 FTFPRPMTDRPGLDPSFGSLKTFPAANTIRANGDD-----QTRADIARAFEDAVDT 246

QY 122 LVKRTHRALPCKQDRLPQNNNAVVASGVASNFYIRALBITLNATQCTLLCPPRL 181
DB 247 LMKCKRAL-----DLQPKR--LVMAAGVSANRLTKAKLAEMMKRGSGVFPARBP 298

```

QY      182 TDNGIMAMNGIERLRAGCIIHLDIEGIRYREKPCPL 217
          ||| ||| : : | : : | : : ||
Db      299 TDNGAMIAVAGMVFKA--GATADL-GVSVRPRWPL 331

```

Search completed: November 10, 2005, 16:53:09
Job time : 15.8076 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2005, 15:16:01 ; Search time 57.9315 Seconds
(without alignments)
2112.614 Million cell updates/sec

Title: US-10-649-273-2_COPY_176_414

Perfect score: 1240

Sequence: 1 LLLVGVSDFLILGKSLDI.....DISKEVGASIKVQLKKEI 239

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt_03:*

1: uniprot_sprot:*

2: uniprot_trembl:*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result No. Score Query Match Length DB ID Description

1	1240	100.0	414	2	Q96EV9	Q96EV9 homo sapien
2	1213.5	97.9	439	2	Q9H4B0	Q9H4B0 homo sapien
3	1086	87.6	414	2	Q6PEB4	Q6PEB4 mus musculu
4	1079	87.0	414	2	Q8BLB6	Q8BLB6 mus musculu
5	1071	86.4	414	2	Q9D0N0	Q9D0N0 mus musculu
6	983	79.3	467	2	Q6AYN7	Q6AYN7 rattus norv
7	960	77.4	364	2	Q96NH5	Q96NH5 homo sapien
8	751	60.6	404	2	Q8JFW3	Q8JFW3 brachydanio
9	747	60.2	404	2	Q8JFR7	Q8JFR7 brachydanio
10	361.5	29.2	401	2	Q7Q9T8	Q7Q9T8 anopheles g
11	332.5	26.8	323	2	Q9VMD6	Q9VMD6 drosophila
12	324	26.1	480	2	Q22145	Q22145 arbidopsis
13	299.5	24.2	255	2	Q677H2	Q677H2 hyacinthus
14	262.5	21.2	360	2	Q92LH8	Q92LH8 rhizobium m
15	257	20.7	335	2	Q7VXN4	Q7VXN4 bordetella
16	254	20.5	346	2	Q73H71	Q73H71 wolbachia p
17	253	20.4	362	2	Q98B16	Q98B16 rhizobium l
18	252.5	20.4	340	2	Q6FCR9	Q6FCR9 actinobact
19	252	20.3	346	2	Q7M6E8	Q7M6E8 bordetella
20	251	20.2	346	2	Q7M134	Q7M134 bordetella
21	246.5	19.9	389	2	Q6NDX3	Q6NDX3 rhodospseudo
22	244	19.8	387	2	Q68XR3	Q68XR3 rickettsia
23	244	19.7	365	2	Q8UC47	Q8UC47 agrobacteri
24	244	19.5	359	2	Q8YJBI	Q8YJBI bruceella me
25	241.5	19.5	359	2	Q8FYI5	Q8FYI5 bruceella su
26	241	19.3	251	2	Q93FU2	Q93FU2 cowdria rum
27	238	19.2	387	1	GCP_RICPR	GCP_RICPR rickettsia
28	237.5	19.2	367	2	Q9ABZ9	Q9ABZ9 caulobacter

32	236.5	19.1	357	2	Q89WM1	Q89WM1 bradyrhizob
33	234.5	18.9	344	2	Q7PAG7	Q7PAG7 rickettsia
34	233.5	18.8	339	2	Q6LV10	Q6LV10 photobacter
35	231	18.6	341	2	Q7NUE3	Q7NUE3 chromobacter
36	231	18.6	341	2	Q91SV7	Q91SV7 pseudomonas
37	231	18.6	364	2	Q6G1R3	Q6G1R3 bartonella
38	230.5	18.6	344	2	Q9ZAK6	Q9ZAK6 rickettsia
39	228.5	18.4	353	2	Q7VQO9	Q7VQO9 candidatus
40	224	18.1	337	2	Q82XN2	Q82XN2 nitrosomona
41	224	18.1	342	1	GCP_HABIN	P43764 haemophilus
42	223.5	18.0	344	2	Q65BP0	Q65BP0 manheimia
43	221.5	17.9	323	2	Q291S3	Q291S3 archaeglob
44	219.5	17.7	347	2	Q8XX97	Q8XX97 ralestonia s
45	217	17.5	341	2	Q8BS16	Q8BS16 oceanobactl

ALIGNMENTS

RESULT 1
Q96EV9 PRELIMINARY: PRT; 414 AA.
ID Q96EV9
AC Q96EV9; 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE OSGBPL1 protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RX MEDLINE=22388257; PubMed12477932; DOI=10.1073/pnas.242603899;
RA Strassberg R.L., Feingold E.A., Grouse L.H., Derye J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schaller G.D.,
RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heish F.,
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abremson R.D., Mullany S.J.,
RA Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richardson S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
FAhey J., Helton B., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Bouffard G.G.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Scherch A., Schein J.E.,
RA Jones S.J., Marra W.A.,
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RA Strassberg R.L.
RT Submitted (JUL-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; BC011904; AAH1904.1; -
DR MEROPS; M22.004; -
DR GO; GO:0008450; F-0-sialoglycoprotein endopeptidase activity; IEA.
DR GO; GO:0008270; F-izing ion binding; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009095; Peptidase_M22.
DR InterPro; IPR009180; Pept_M22_Osialgl.
DR Pfam; PF00814; Peptidase_M22_1.
DR PIRSF; PIRSF004537; Osialglc_ptide; 1.
DR PRINTS; PR00789; OSIALOPTASE.
DR Prodom; PD002367; Peptidase_M22; 1.
DR TIGRFAMs; TIGR00329; GCP_1_
SQ SEQUENCE 414 AA; 45122 MW; A536B333F5C68BD CRC64;

Query Match 100.0%; Score 1240; DB 2; Length 414;
 Best Local Similarity 100.0%; Pred. No. 5,6e-101;
 Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLAIVGVSDPFLILGKSLDIAPGDMLDKVARRLSLIKHEPCSTMSGKAIIEHLAKQGNRF 60
 DB 176 LLAIVGVSDPFLILGKSLDIAPGDMLDKVARRLSLIKHEPCSTMSGKAIIEHLAKQGNRF 235
 QY 61 HFDIKPPLHAAKNCDFSTGLOHVTDKIIMKKEKEGIEKGQILSSAADIAATVOHTMAC 120
 DB 236 HFDIKPPLHAAKNCDFSTGLOHVTDKIIMKKEKEGIEKGQILSSAADIAATVOHTMAC 295
 QY 121 HLAVKTRHAILFCCKORDLLPQNNAVLVASGVASFYIRRALEILTNAQTCTLLCPPRL 180
 DB 296 HLAVKTRHAILFCCKORDLLPQNNAVLVASGVASFYIRRALEILTNAQTCTLLCPPRL 355
 QY 181 CTGNGIMIANWGIERLRAGLIHDIEGIRPEKCPGLGVDSKEVGEASIKVPOLKMEI 239
 DB 356 CTGNGIMIANWGIERLRAGLIHDIEGIRPEKCPGLGVDSKEVGEASIKVPOLKMEI 414

RESULT 2

Q9H4B0 PRELIMINARY; PRT; 439 AA.
 ID 09H4B0;
 AC 09H4B0;
 DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
 DE Putative sialoglycoprotease type 2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NC NCB1_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Uterus;
 RA Chen J.M., Fortunato M., Barrett A.J.;
 RL Submitted (OCP-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AJ295148; CAC14666.1; --
 DR MEROPS; M22.004; --
 DR Genew; HGNC:23075; OSGEPL1.
 DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.
 DR GO; GO:0008433; F:peptidase activity; IEA.
 DR GO; GO:0008270; F:zinc ion binding; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000905; Peptidase_M22.
 DR InterPro; IPR009180; Pept_M22_Osialyl.
 DR Pfam; PF00814; Peptidase_M22; 1.
 DR PIRSF; PIRSF004537; Osialylc_pptds; 1.
 DR PRINTS; PR00789; OSIALOPTASE.
 DR Prodom; PD002367; Peptidase_M22; 1.
 DR TrifPfam; TIGR00329; gcp; 1.
 KW Protease.
 SQ SEQUENCE 439 AA; 48040 MW; 44849372C784E41F CRC64;

Query Match 97.8%; Score 1213.5; DB 2; Length 439;
 Best Local Similarity 90.2%; Pred. No. 1.3e-98;
 Matches 238; Conservative 0; Mismatches 1; Indels 25; Gaps 1;

QY 1 LLAIVGVSDPFLILGKSLDIAPGDMLDKVARRLSLIKHEPCSTMSGKAIIEHLAKQGNRF 60
 DB 176 LLAIVGVSDPFLILGKSLDIAPGDMLDKVARRLSLIKHEPCSTMSGKAIIEHLAKQGNRF 235
 QY 61 HFDIKPPLHAAKNCDFSTGLOHVTDKIIMKKEKEGIEKGQILSSAADIAATVOHTMAC 98
 DB 236 HFDIKPPLHAAKNCDFSTGLOHVTDKIIMKKEKEGIEKGQILSSAADIAATVOHTMAC 295
 QY 99 ---EKGQILSSAADIAATVOHTMACHLVKTTRAILFCCKORDLLPQNNAVLVASGVASN 155
 DB 296 CRVEKQILSSAADIAATVOHTMACHLVKTTRAILFCCKORDLLPQNNAVLVASGVASN 355
 QY 156 FYIRALBELITNAQTCTLLCPPRLCTGNGIMIANWGIERLRAAGLIHDIEGIRPEK 215

DB 356 FYIRALBELITNAQTCTLLCPPRLCTGNGIMIANWGIERLRAAGLIHDIEGIRPEK 415
 QY 216 PLGVDSKEVGEASIKVPOLKMEI 239
 DB 416 PLGVDSKEVGEASIKVPOLKMEI 439

RESULT 3

OSPEB4 PRELIMINARY; PRT; 414 AA.
 ID 06PEB4;
 AC 06PEB4;
 DT 05-JUL-2004 (TREMBlrel. 27, Created)
 DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
 DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
 DE Hypothetical protein.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NC NCB1_TaxID=10090;
 RN [1]

SEQUENCE FROM N.A.

RP STRAIN=CZECH II; TISSUE=Mammary tumor;
 RC MEDLINE=22389257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RX Strausberg R.L., Pelngold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shemmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heish F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo W.F., Casavant J.L., Schetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McMan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahney J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalska U., Smalls D.E., Schmeich A., Schein J.E.,
 RA Jones S.J., Matra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]

SEQUENCE FROM N.A.

RP STRAIN=CZECH II; TISSUE=Mammary tumor;
 RC Strausberg R.;
 RA Submitted (SEP-2003) to the EMBL/Genbank/DBJ databases.
 RL EMBL; BC058172; AAH58172.1; --
 DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.
 DR GO; GO:0008270; F:zinc ion binding; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000905; Peptidase_M22.
 DR InterPro; IPR009180; Pept_M22_Osialyl.
 DR Pfam; PF00814; Peptidase_M22; 1.
 DR PIRSF; PIRSF004537; Osialylc_pptds; 1.
 DR PRINTS; PR00789; OSIALOPTASE.
 DR Prodom; PD002367; Peptidase_M22; 1.
 DR TrifPfam; TIGR00329; gcp; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 414 AA; 44962 MW; 0418AD93C6FEBCAE CRC64;

Query Match 87.8%; Score 1086; DB 2; Length 414;
 Best Local Similarity 85.8%; Pred. No. 2.2e-87;
 Matches 205; Conservative 15; Mismatches 19; Indels 0; Gaps 0;

QY 1 LLAIVGVSDPFLILGKSLDIAPGDMLDKVARRLSLIKHEPCSTMSGKAIIEHLAKQGNRF 60
 DB 176 LLAIVGVSDPFLILGKSLDIAPGDMLDKVARRLSLIKHEPCSTMSGKAIIEHLAKQGNRF 235
 QY 61 HFDIKPPLHAAKNCDFSTGLOHVTDKIIMKKEKEGIEKGQILSSAADIAATVOHTMAC 120

Db 236 HFTINPQWNAKNCDFSTGLQHTDKLITHEKEGIEKQILSSAADIAAVQHTAC 295
 Qy 121 HLKXTRHAILFCCKQDPLPQNNAVLVASGVASNFYIRALEILTNAQCTLLCPPEL 180
 Db 296 HLAKRTHAILFCCKQNLSPANAVALVSGVASNLYIRKALEIVANATQCTLLCPPEL 355
 Qy 181 CTDNGIMAMNGIERLRAGILHDIEGIRYEPKCPGLGVDISKEVGEASIKVPOLKMEI 239
 Db 356 CTDNGIMAMNGIERLRAGILHDIEGIRYEPKCPGLGVDISKEVGEASIKVPOLKMEI 414
 RESULT 4
 Q8BLB6 PRELIMINARY; PRT; 414 AA.
 AC Q8BLB6; 1
 DT 01-MAR-2003 (TREMBLrel. 23, Created)
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
 DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
 DE Mus musculus adult male corpora quadrigemina cDNA, RIKEN full-length
 DE enriched library, clone:B230219017 product:similar to PUTATIVE
 DE SIALOGLYCOPROTEASE TYPE 2.
 GN Name=Osgp11;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;
 RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
 RA Carninci P., Hayashizaki Y.;
 RT "High-efficiency full-length cDNA cloning.";
 RL Meth. Enzymol. 303:19-44(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;
 RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
 RA RIKEN PANTOM Consortium;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;
 RA The PANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs.";
 RL Nature 420:563-573(2002).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;
 RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
 RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to
 RT prepare full-length cDNA libraries for rapid discovery of new genes.";
 RL Genome Res. 10:1617-1630(2000).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;
 RX MEDLINE=70530913; PubMed=11076861; DOI=10.1101/gr.152600;
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
 RA Kono H., Akiyama J., Nishi K., Katsunai T., Tashiro H., Itoh M.,
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishino T., Harada A.,
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
 RA Fujisake S., Inoue K., Togawa Y., Izawa M., Ohara B., Watabiki M.,
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-Format
 RT sequencing pipeline with 384 multichipillary sequencer.";
 RL Genome Res. 10:1757-1771(2000).
 RN [6]
 RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;
 RA Aachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
 RA Fukuda S., Furuta M., Hanagaki T., Hara A., Hashizume W.,
 RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
 RA Horii F., Imotani K., Ishii Y., Itoh M., Kaga I., Kasubawa T.,
 RA Kato H., Kawai J., Kojima Y., Kondo S., Kono H., Kouda M., Koya S.,
 RA Kurihara C., Matsuyama T., Miyazaki A., Murata N., Nakamura M.,
 RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
 RA Saito K., Saitoh K., Sakai C., Sakai C., Sakazume N., Sano H.,
 RA Sasaki D., Shibata K., Shingawa A., Shiraki T., Sogabe Y., Tagami M.,
 RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
 RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBD databases.
 DR EMBL; AK045669; BAC2450.1; -
 DR MEROPS; M22.004; -
 DR MGD; MG11919335; Osgp11.
 DR GO; GO:0008450; P.O.-sialoglycoprotein endopeptidase activity; IEA.
 DR GO; GO:0008233; Fipeptidase activity; IEA.
 DR GO; GO:0008270; F:zinc ion binding; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000905; Peptidase M22.
 DR InterPro; IPR009180; Pept_M22_Osialgl.
 DR Pfam; PF00814; Peptidase_M22_1.
 DR PRINTS; PRS004537; Osialglc_ptcls; 1.
 DR PRODOM; PD002367; OsialOPTASE.
 DR TIGRPFAM; TIGR00329; gcp; 1.
 KM Protease.
 SQ SEQUENCE 414 AA; 44927 MW; 9A2D27B576C7DB24 CRC64;
 Query Match 87.0%; Score 1079; DB 2; Length 414;
 Best Local Similarity 85.8%; Pred. No. 9.1e-87;
 Matches 205; Conservative 14; Mismatches 20; Indels 0; Gaps 0;
 Qy 1 LALVGVSDPFLIGSLDIPAGDMIDKVARSLILKHPFCSTMGSKALEIAKQGNRF 60
 Db 176 LIALVGVSDPFLIGSLDIPAGDMIDKVARSLILKHPFCSTMGSKALEIAKQGNRF 235
 Qy 61 HPIKREPLHAKNCDFSTGLQHTDKLITHEKEGIEKQILSSAADIAAVQHTAC 120
 Db 236 HFTINPQWNAKNCDFSTGLQHTDKLITHEKEGIEKQILSSAADIAAVQHTAC 295
 Qy 121 HLKXTRHAILFCCKQDPLPQNNAVLVASGVASNFYIRALEILTNAQCTLLCPPEL 180
 Db 296 HLAKRTHAILFCCKQNLSPANAVALVSGVASNLYIRKALEIVANATQCTLLCPPEL 355
 Qy 181 CTDNGIMAMNGIERLRAGILHDIEGIRYEPKCPGLGVDISKEVGEASIKVPOLKMEI 239
 Db 356 CTDNGIMAMNGIERLRAGILHDIEGIRYEPKCPGLGVDISKEVGEASIKVPOLKMEI 414
 RESULT 5
 Q9DONO PRELIMINARY; PRT; 414 AA.
 AC Q9DONO; 1
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
 DE Mus musculus 10 days embryo whole body cDNA, RIKEN full-length
 DE enriched library, clone:2610001M19 product:similar to PUTATIVE
 DE SIALOGLYCOPROTEASE TYPE 2.
 GN Name=Osgp11;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=whole body;
 RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
 RA Carninci P., Hayashizaki Y.;
 RT "High-efficiency full-length cDNA cloning.";
 RL Meth. Enzymol. 303:19-44(1999).

RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Whole body;
 RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
 RA RIKEN FANTOM Consortium;
 RT "Functional annotation of a full-length mouse cDNA collection."
 RL Nature 409:685-690(2001).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Whole body;
 RA The FANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs."
 RL Nature 420:563-573(2002).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Whole body;
 RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
 RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to
 RT prepare full-length cDNA libraries for rapid discovery of new genes."
 RL Genome Res. 10:1617-1630(2000).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Whole body;
 RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
 RA Kono H., Akiyama J., Nishi K., Katsunai T., Taishiro H., Itoh M.,
 RA Suni N., Ishii Y., Nakamura S., Hazama M., Nishino T., Harada A.,
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashigaki K.,
 RA Fujiwaki S., Inoue K., Togawa Y., Izawa M., Ohara E., Wachihi M.,
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsunura S., Kawai J.,
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format
 RT sequencing pipeline with 384 multicapillary sequencer."
 RL Genome Res. 10:1757-1771(2000).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Whole body;
 RA Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
 RA Arkawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
 RA Haraagaki T., Hara A., Hayatsu N., Hiramoto K., Hirooka T., Hori F.,
 RA Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
 RA Kawai J., Kojima Y., Kono H., Konda M., Koya S., Kurihara C.,
 RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Nunazaki R., Ohno M.,
 RA Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,
 RA Sano H., Sasaki D., Shibata K., Shibata Y., Shingawa S., Shiraki T.,
 RA Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
 RA Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
 RA Muramatsu M., Hayashizaki Y.;
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 RN EMBL: AK011265; BAB27506.1; -
 DR MEROPS: M22.004; -
 DR MGI: 1919335; Osgsep11.
 DR GO: GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.
 DR GO: GO:0008233; F:peptidase activity; IEA.
 DR GO: GO:0008270; F:zinc ion binding; IEA.
 DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro: IPR000905; Peptidase_M22.
 DR InterPro: IPR009180; Pept_M22_Osialgl.
 DR Pfam: PF00814; Peptidase_M22; 1.
 DR PIRSF: PIRSF004537; Osialglc_ptcds; 1.
 DR PRINTS: PR00789; OSIALOPTASE.
 DR PRODOM: PD002367; Peptidase_M22; 1.
 DR TIGRFAMs: TIGR00329; gcp; 1.
 DR TrEMBL: -
 KW PROTEASE.
 SQ SEQUENCE 414 AA; 44939 MW; 999BC689944DDB24 CRC64;

QY 1 LIALVGVSDPFLILKSLIDIAFGDMLDKVARRLSILKHBCSTMSGKAIEHLAKGNRF 60
 DB 176 LIALVGVSDPFLILKSLIDIAFGDMLDKVARRLSILKHBCSTMSGKAIEHLAKGNRF 235
 QY 61 HFDIYKPLHLAAKCPDFSPFLGQVHTDKIMKEKEKEGIRKQGLSSAAIDIAVQHTMVC 120
 DB 236 HFTINPMPQNAKQCFSPFLGQVHTDKIMKEKEKEGIRKQGLSSAAIDIAVQHTMVC 295
 QY 121 HIKKTRRALPCKORDLLPQNNAVLVASGVASNYIRALAEILTNATQCTLLCPPLR 180
 DB 296 HIAKTRHRLPCKQKNLSPNAAVLVSGVASNYIRKALEIVANAQCTLLCPPLR 355
 QY 181 CTDNGIMIMANGIERIRAGIGIHDIEGIRYBPKPLGYDISKEVBSAIXVQLKMEI 239
 DB 356 CTDNGIMIMANGIERIRAGIGIHDIEGIRYBPKPLGYDISKEVBSAIXVQLKMEI 414
 RESULT 6
 O6AVN7 PRELIMINARY; PRT; 467 AA.
 ID O6AVN7;
 AC 25-OCT-2004 (TrEMBLrel. 28, Created)
 DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
 DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
 DE Hypothetical protein (Fragment).
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_Taxid=10116;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RX PubMed=12477932; DOI=10.1073/pnas.242603999;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,
 RA Dietzenko L., Marsina K., Farmer A.A., Rubin G.W., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Umed T.B., Toshlyuk S., Carninci P., Prange C.,
 RA Raha S.S., Loughlano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulys S.W.,
 RA Villalon D.K., Murry D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butlerfield Y.S.,
 RA Krzywiński M.I., Skalka U., Smalms D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RA Director MGC Project;
 RL EMBL: BC078974; AAH78974.1; -
 DR GO: GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.
 DR GO: GO:0008270; F:zinc ion binding; IEA.
 DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro: IPR000905; Peptidase_M22.
 DR InterPro: IPR009180; Pept_M22_Osialgl.
 DR InterPro: IPR002016; Peroxidase.
 DR Pfam: PF00814; Peptidase_M22; 1.
 DR PIRSF: PIRSF004537; Osialglc_ptcds; 1.
 DR PRINTS: PR00789; OSIALOPTASE.
 DR PRODOM: PD002367; Peptidase_M22; 1.
 DR TIGRFAMs: TIGR00329; gcp; 1.
 DR PROSITE: PS00435; PEROXIDASE_1; UNKNOWN_1.
 KW Hypothetical protein.

Query Match 86.4%; Score 1071; DB 2; Length 414;
 Best Local Similarity 85.4%; Pred. No. 4.6e-86;
 Matches 204; Conservative 14; Mismatches 21; Indels 0; Gaps 0;

FT NON TER 467 467
 SQ SEQUENCE 467 AA; 50799 MW; 474E1B1959B9AC0 CRC64;
 Query Match 79.3%; Score 983; DB 2; Length 467;
 Best Local Similarity 87.9%; Pred. No. 3.1e-78;
 Matches 186; Conservative 9; Mismatches 17; Indels 0; Gaps 0;

QY 1 LIALVQGVSDPFLILGKSLDIAPGMDLKVARRLSLIKHEPCSTMSGGKAIEHLAKQGNRF 60
 DB 176 LIALVQGVSDPFLILGKSLDIAPGMDLKVARRLSLIKHEPCSTMSGGKAIEHLAKQGNRF 235
 QY 61 HFDIKPPLHAKNCDSPFTGLQHTVDKLIIMKKEKEGIEKGQILSSAADIAATVQHTMAC 120
 DB 236 HFTINPPEQNAKNCDFSTGLQHTVDKLIIMKKEKEGIEKGQILSSAADIAATVQHTMAC 295
 QY 121 HLKRTTRAILFCCKORDLIPONNAVIVASGVANFYIRRALEITLNATQCTLLCPPRL 180
 DB 296 HLKRTTRAILFCCKORDLIPONNAVIVASGVANFYIRRALEITLNATQCTLLCPPRL 355
 QY 181 CTDNGIMIAWNGIERLRAGLGIHDIIEGIRYEPK 214
 DB 356 CTDNGIMIAWNGIERLRAGLGIHDIIEGIRYEPK 389

RESULT 7
 Q96NH5 PRELIMINARY; PRT; 364 AA.
 AC Q96NH5;
 DT 01-DEC-2001 (TREMblrel. 19, Created)
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
 DT 01-MAR-2004 (TREMblrel. 26, Last annotation update)
 DE Hypothetical protein FLJ30879.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX PubMed=14702039; DOI=10.1038/ng1285;
 RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
 RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
 RA Sekine M., Ohyashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
 RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahara K.,
 RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,
 RA Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
 RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,
 RA Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,
 RA Niinomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,
 RA Tanai H., Kimata M., Watanabe S., Hiraoa S., Chiba Y., Ishida S.,
 RA Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J.,
 RA Kanehori K., Takahashi-Fuji A., Hara H., Tanase T., Nomura Y.,
 RA Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,
 RA Nishizawa Y., Matsunawa H., Ichihara T., Shiohara T., Sano S.,
 RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohara T., Sano S.,
 RA Moriya S., Momiya H., Satoh N., Takami S., Teraishi Y., Suzuki O.,
 RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
 RA Hishiyaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
 RA Yamazaki H., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
 RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujimori T.,
 RA Ono T., Yamada K., Fujii Y., Ozaki K., Hiroo M., Omori Y.,
 RA Okabe A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
 RA Ohtsuri R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
 RA Matsunura K., Nakajima Y., Mizuno T., Morinaga M., Saeki M.,
 RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
 RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
 RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuno Y., Yamashita R.,
 RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;
 RT *Complete sequencing and characterization of 21,243 full-length human
 RT cDNA's;
 RT Nat. Genet. 36:40-45(2004).
 RL EMBL; AK055441; BAB70923.1; -
 DR MEROPS; M22.004; -

DR GO; GO:0008450; F.O-sialoglycoprotein endopeptidase activity; IEA.
 DR GO; GO:0008233; F:peptidase activity; IEA.
 DR GO; GO:0008270; F:zinc ion binding; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000905; Peptidase M22.
 DR InterPro; IPR009180; Pept_M22_Osialgl.
 DR Pfam; PF00814; Peptidase_M22; 1.
 DR PIRSF; PIRSF04537; Osialglc_pptds; 1.
 DR PRINTS; PR00789; OSIALOPTASE.
 DR ProDom; PD002367; Peptidase_M22; 1.
 DR TIGRFAMs; TIGR00329; gcp; 1.
 DR TrEMBL; TrEMBL:Q96NH5; 1.
 SQ SEQUENCE 364 AA; 39528 MW; E0E605A07D0EC3D6 CRC64;
 Query Match 77.4%; Score 960; DB 2; Length 364;
 Best Local Similarity 98.4%; Pred. No. 2.5e-76;
 Matches 186; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 LIALVQGVSDPFLILGKSLDIAPGMDLKVARRLSLIKHEPCSTMSGGKAIEHLAKQGNRF 60
 DB 176 LIALVQGVSDPFLILGKSLDIAPGMDLKVARRLSLIKHEPCSTMSGGKAIEHLAKQGNRF 235
 QY 61 HFDIKPPLHAKNCDSPFTGLQHTVDKLIIMKKEKEGIEKGQILSSAADIAATVQHTMAC 120
 DB 236 HFDIKPPLHAKNCDSPFTGLQHTVDKLIIMKKEKEGIEKGQILSSAADIAATVQHTMAC 295
 QY 121 HLKRTTRAILFCCKORDLIPONNAVIVASGVANFYIRRALEITLNATQCTLLCPPRL 180
 DB 296 HLKRTTRAILFCCKORDLIPONNAVIVASGVANFYIRRALEITLNATQCTLLCPPRL 355
 QY 181 CTDNGIMIA 189
 DB 356 CTDNGIMIA 364

RESULT 8
 Q96FW3 PRELIMINARY; PRT; 404 AA.
 AC Q96FW3;
 DT 01-OCT-2002 (TREMblrel. 22, Created)
 DT 01-OCT-2002 (TREMblrel. 22, Last sequence update)
 DT 01-MAR-2004 (TREMblrel. 26, Last annotation update)
 DE SI:d2211013.4 (Novel glycoprotease).
 GN Name=d2211013.4;
 OS Brachydanio rerio (Zebrafish) (Danio rerio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Danio.
 NC NCBI_TaxID=7955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Babage A.;
 RL Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.
 RL EMBL; AL591593; CND43471.1; -
 DR MEROPS; M22.004; -
 DR GO; GO:0008450; F.O-sialoglycoprotein endopeptidase activity; IEA.
 DR GO; GO:0008233; F:peptidase activity; IEA.
 DR GO; GO:0008270; F:zinc ion binding; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000905; Peptidase M22.
 DR InterPro; IPR009180; Pept_M22_Osialgl.
 DR Pfam; PF00814; Peptidase_M22; 1.
 DR PIRSF; PIRSF04537; Osialglc_pptds; 1.
 DR PRINTS; PR00789; OSIALOPTASE.
 DR ProDom; PD002367; Peptidase_M22; 1.
 DR TIGRFAMs; TIGR00329; gcp; 1.
 DR TrEMBL; TrEMBL:Q96FW3; 1.
 SQ SEQUENCE 404 AA; 43956 MW; 3A6B1DD050737B35 CRC64;
 Query Match 60.6%; Score 751; DB 2; Length 404;
 Best Local Similarity 61.6%; Pred. No. 7.9e-58;
 Matches 146; Conservative 32; Mismatches 59; Indels 0; Gaps 0;

RA [1]
RA SEQUENCE FROM N.A.
RA STRAIN=Berkley;
RA Stapleton M., Brokstein P., Hong L., Agpayani A., Carlson J.,
RA Champagne M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Gurin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nuno J., Pacleb J., Paragas V., Park S., Pounanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.,
RA Submitted (Aug-2001) to the EMBL/Genbank/DBJ databases.
RA EMBL: AY051882; AAK93306.1; -
RA FlyBase; FBgn031060; CG14231.
RA GO: GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.
RA GO: GO:0008270; F:zinc ion binding; IEA.
RA GO: GO:0006508; P:proteolysis and peptidolysis; IEA.
RA InterPro; IPR009180; Pept M22_Osialgl.
RA Pfam; PF00814; Peptidase M22; 1.
RA PIRSF; PIRSF004537; Osialglc.ptids; 1.
RA PRINTS; PR00789; OSIALOPTASE.
RA SEQUENCE 323 AA; 35828 MW; BEDB854D9A2BF35A CRC64;
SQ
Query Match 26.8%; Score 332.5; DB 2; Length 323;
Best Local Similarity 35.9%; Pred. No. 5.3e-21;
Matches 83; Conservative 38; Mismatches 97; Indels 13; Gaps 5;
QY 2 LAIVQGVSDFLKSLDIAPGMDLKVARRLSLKHEPCTSGSKAIEHLAK-QGNRF 60
DB 81 LVVANGPGRLLTLGQLDDADGAEAFDKIGRLRLHLLEFRLMNGRAIEHNAQLASPL 140
QY 61 HFDIKPLHAKNCDFFSTGLQHTYDKIMKKEKEEGEIGKQIASSADIATVQHTMAC 120
DB 141 AYEPLPLAQQRCNCFSPAGIKNNFRAIRARERPPDGVISNYDPCFAGLRSYSR 200
QY 121 HLVEKTRATLFC--KORDLLPQNNAVLVAGVANSFYIRRAELTNATQCTLLCP 178
DB 201 HAKMTKRALEICLLPHRQLFGDTPPTLVMSGGVANNALITANIEHLAQICRSFRPSK 260
QY 179 RLCTDNGIMIAMNGIERLRAGILHDIIE-GIRYEPKCPGLDISKEVGEA 228
DB 261 RYCSDCNVMITAMHGEVL-----LQDKEASTRYDYD---SIDIGSAGFA 302
RESULT 12
Q9VWD6 PRELIMINARY; PRT; 409 AA.
ID Q9VWD6
AC Q9VWD6
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE CG14231-PA.
GN ORFNames=CG14231;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Peerygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RA [1]
RA SEQUENCE FROM N.A.
RA MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gaber G.L.,
RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballou R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burks C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibbegam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuang D., Lai Z.,
RA Lascko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merulov G., Mlshina N.V., Moadary C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reihert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier B., Spreading A.C., Stapleton M., Strong R., Sun E.,
RA Svrtkac R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wasearman D.A., Weinstein G.M., Weissendbach J.,
RA Williams S.M., Woodagel, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of Drosophila melanogaster";
RA Science 287:2185-2195(2000).
RL [12]
RL SEQUENCE FROM N.A.
RL MEDLINE=22426065; PubMed=12537568;
RA Celniker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svrtkac R., Taber P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RA "Finishing a whole-genome shotgun: Release 3 of the Drosophila
RA melanogaster euchromatic genome sequence."
RA Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RL [13]
RL SEQUENCE FROM N.A.
RL MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J., Svrtkac R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celniker S.E.;
RA "The transposable elements of the Drosophila melanogaster euchromatin:
RA a genomic perspective."
RA Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
RL [14]
RL SEQUENCE FROM N.A.
RL MEDLINE=22426069; PubMed=12537572;
RA Mitra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.B.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Betencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.O.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RA "Annotation of the Drosophila melanogaster euchromatic genome: a
RA systematic review."
RA Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
RL [15]
RL SEQUENCE FROM N.A.
RL FlyBase;
RL Submitted (SEP-2002) to the EMBL/Genbank/DBJ databases.
RL [16]
RL SEQUENCE FROM N.A.
RL FlyBase;
RL Submitted (MAR-2004) to the EMBL/Genbank/DBJ databases.
RA EMBL: AE003513; AAF49008.1; -
RA InChit; Q9VWD6; -
RA FlyBase; FBgn031060; CG14231.
DR GO: GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.
DR GO: GO:0008270; F:zinc ion binding; IEA.
DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009095; Peptidase_M22.

DR InterPro: IPR009180; Pept_M22_Osialgl.
 DR Pfam: PF00814; Peptidase_M22; 1.
 DR PIRSF: PIRSF004537; Osialglc_ptids; 1.
 DR PRINTS: PR00789; OSIALOPTASE
 DR PRODOM: PD002367; Peptidase_M22; 1.
 DR TIGRFAMs: TIGR00329; gcp; 1.
 DR SEQUENCE 409 AA; 45328 MW; 97976667D155538 CRC64;
 SQ
 Query Match 26.8%; Score 332.5; DB 2; Length 409;
 Best Local Similarity 35.9%; Pred. No. 6.9e-21;
 Matches 83; Conservative 38; Mismatches 97; Indels 13; Gaps 5;
 QY 2 LATVQGVSDFLILKSLDIPADMDKVARRLSLIKPECSMTSGGKAIEHLAK-QGNRF 60
 DB 167 LVANVGGRLLTGLQTDADAGEAFDKIGRLRLHLEPYRLMNGRAIEHNAQLASDPL 226
 QY 61 HFDIKPPLHAKNCDFSTGLOHTVDKTIIMKKEKEGKGLISLADIAATVQHTMAC 120
 DB 227 AYEPLPLAQQRNCFNCFAGIKNNFRAIRARERERPPDGVISNYGDFCGLRSVS 286
 QY 121 HLVRTRRLIFC--KORDLLPNNNAVIVASGVASNFYIRRALIELTNATQCTLLCPP 178
 DB 287 HLMHRTPLALEYCLLPHRQLFGDTPPTLVMSGVANNDAIYANIHLLAQIGCRSFRPSK 346
 QY 179 RLCTDNGIMIAMNGIERLRAGILHDIE-GIRYEPKCPGLGVDSKEVGEA 228
 DB 347 RYCSDNGVMIMHVEQL-----LQDKKASTRYDYD---SIDIGSGNFA 388
 RESULT 13
 AC 022145 PRELIMINARY; PRT; 480 AA.
 ID 022145; OSVWL2;
 DT 01-JAN-1998 (TRENBLrel. 05, Created)
 DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)
 DT 05-JUL-2004 (TRENBLrel. 27, Last annotation update)
 DE Putative O-sialoglycoprotein endopeptidase (Sialoglycoprotease GCP1).
 GN Name=At2g45270; Synonyms=GCP1;
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OC NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Rounsley S.D., Lin X., Ketchum K.A., Crosby M.L., Brandon R.C.,
 RA Sykes S.M., Mason T.M., Kerlavage A.R., Adams M.D., Somerville C.R.,
 RA Venter J.C.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Town C.D., Kaul S.;
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Hausseuhl K., Adamska I.;
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
 RA Goldsmith A.D., Lee J.M., Quach H.L., Toriumi M., Yu G., Bowser L.,
 RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
 RA Kamiya A., Karlin-Neumann G., Kawai U., Kim C., Lam B., Lin J.,
 RA Kamiya A., Karlin-Neumann G., Kawai U., Kim C., Lam B., Lin J.,
 RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,
 RA Seki M., Shim P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
 RA Theologis A.;
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Banh J., Chan M.M., Chang C.H., Chang E., Dale J.M.,
 RA Deng J.M., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.L.,
 RA Tang C., Toriumi M., Wu H.C., Yamamura Y., Yu G., Bowser L.,

RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
 RA Kamiya A., Karlin-Neumann G., Kawai U., Kim C., Lam B., Lin J.,
 RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,
 RA Seki M., Shim P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
 RA Theologis A.;
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC002387; AAB82636.2; -
 DR EMBL: AY024338; AAK00530.1; -
 DR EMBL: AY063864; AAL36220.1; -
 DR EMBL: AY117283; AAM51358.1; -
 DR PIR: E84888; E84888.
 DR GO: GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.
 DR GO: GO:0008233; F:peptidase activity; IEA.
 DR GO: GO:0008270; F:zinc ion binding; IEA.
 DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro: IPR009095; Peptidase_M22
 DR InterPro: IPR009180; Pept_M22_Osialgl.
 DR Pfam: PF00814; Peptidase_M22; 1.
 DR PIRSF: PIRSF004537; Osialglc_ptids; 1.
 DR PRINTS: PR00789; OSIALOPTASE.
 DR PRODOM: PD002367; Peptidase_M22; 1.
 DR TIGRFAMs: TIGR00329; gcp; 1.
 KW Protease.
 SQ SEQUENCE 480 AA; 52995 MW; 20DD6A86ACCFAD CRC64;
 Query Match 26.1%; Score 324; DB 2; Length 480;
 Best Local Similarity 34.1%; Pred. No. 4.7e-20;
 Matches 86; Conservative 34; Mismatches 86; Indels 46; Gaps 6;
 QY 1 LIALVQGVSDFLILKSLDIPADMDKVARRLSLIKPECSMTSGGKAIEHLAKGNRF 60
 DB 224 LILVLAHKLGOYQLGTGVDDAIGEAFTAKVIGLDMH-----RSSGPAVEBELLEGDAK 278
 QY 61 HFDIKPPLHAKNCDFSTGLOHTVDKTIIMKKEKEGKGLISLADIAATVQHTMAC 112
 DB 279 SVKENVPMKYHDCNFSYAGLQYRLALEAE---IDAKCPVSAATNEDRRNADIA 334
 QY 113 TVQHTMACHLVTRTRALIFC--KORDLLPNNNAVIVASGVASNFYIRRALIELTNATQCT 172
 DB 335 SFQRAVLAHLBEKCEKRAIDMALE---LEPSIKMVTSGVASKYRLNLTNVEKNL 391
 QY 173 ILCPPEPLCTDNGIMIAMNGIERLRAGILHDIEGIRY-----PKCP 216
 DB 392 LVCPPEPLCTDNGVMVMTGLTLEHFRVG-----RYDPPPEATEPEDYVYDLRBP 441
 QY 217 LGVDISKEVGEA 228
 DB 442 LGSEYAKGRSEA 453
 RESULT 14
 ID 0677H2 PRELIMINARY; PRT; 255 AA.
 AC 0677H2;
 DT 25-OCT-2004 (TRENBLrel. 28, Created)
 DT 25-OCT-2004 (TRENBLrel. 28, Last sequence update)
 DT 25-OCT-2004 (TRENBLrel. 28, Last annotation update)
 DE O-sialoglycoprotein endopeptidase (Fragment).
 OS Hyacinthus orientalis.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Hyacinthaceae; Magnoliophyta; Liliopsida; Asparagales; Hyacinthaceae;
 OC Hyacinthus.
 OC NCBI_TaxID=82025;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Floral meristem 5-10 days when regenerated in vitro;
 RA Fan J.H., Ma Y., Zhang X.S.;
 RT mRNA, expressed during regeneration of floral bud.
 RL Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY389554; AAT08652.1; -
 DR InterPro: IPR009095; Peptidase_M22.
 DR PRINTS: PR00789; OSIALOPTASE.

DR Prodom: PD002367; Peptidase_M22; 1.
FT NON_TER 1 1
FT NON_TER 255 255
SQ SEQUENCE 255 AA; 28516 MW; 529D2266C284106 CRC64;
Query Match 24.2%; Score 299.5; DB 2; Length 255;
Best Local Similarity 37.4%; Pred. No. 3.3e-18;
Matches 79; Conservative 26; Mismatches 79; Indels 27; Gaps 5;
QY 14 LGSGLDIAPGMDLKVARRSLTIKHPECSMTSGKAIEHLAKQGNRPHFDIPPLHAKN 73
DB 1 LGTTIDDAIGAYDKTARWLG-----DMGKGGGALBELALEGDANSYKFIIPKQKQ 55
QY 74 CDFSEFTGLQHTVDKIIIMKEKEEGIEKQILSSA-----ADIAATVQHTNACHLVKRT 126
DB 56 CNFSYAGLK--TVRLAIESRNICIDKSPISLATCHDRMRADIASFQAVLHLEDR 113
QY 127 HRAILFCQKQRLDLPQNNAVLVASGVASNFYIRRALIELTNATQCTLLCPPRLCTDNGI 186
DB 114 ARAIEMALIEIPTVO---CLVVGSGVASNKYVRSRLNHLVKTGLRLVCPPLCTDNGV 170
QY 187 MIAWNGIERLARGILHDIGIREPCPL 217
DB 171 MWMTGIEENFLVG-----RYDPPPPV 191
RESULT 15
092LH8 PRELIMINARY; PRT; 360 AA.
AC 092LH8;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DE 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE PROBABLE O-SIALOGLYCOPROTEIN ENDOPEPTIDASE (EC 3.4.24.57).
GN ORFNames=SMC03230;
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N. A.
RC STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430; DOI=10.1073/pnas.161294398;
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Barut J.,
RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Godrie T., Goffeau A., Kahn D., Kiss B., Lelaure V., Masuy D.,
RA Pohl T., Portetelle D., Puehler A., Purnelle B., Ramsperger U.,
RA Renard C., Thebaud P., Vandenbol M., Weidner S., Galibert F.;
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021."
RT Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
RL EMBL; AL591792; CAC47657.1; -.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000905; Peptidase_M22.
DR InterPro; IPR009180; Pept_M22_Osialgl.
DR Pfam; PF00814; Peptidase_M22; 1.
DR PIRSF; PIRSF004537; Osialglc_ptide; 1.
DR PRINTS; PR00789; OSIALOPTASE.
DR Prodom; PD002367; Peptidase_M22; 1.
DR TIGRPFMs; TIGR00329; gcp; 1.
KW Complete proteome; Hydrolase.
SQ SEQUENCE 360 AA; 37906 MW; A07F946AB562EA86 CRC64;
Query Match 21.2%; Score 262.5; DB 2; Length 360;
Best Local Similarity 32.8%; Pred. No. 9e-15; Indels 43; Gaps 7;
Matches 78; Conservative 32; Mismatches 85; Indels 43; Gaps 7;
QY 2 LALVQGVSDFLILCKSLDIAPGMDLKVARRSLTIKHPECSMTSGKAIEHLAKQGNRPH 61
DB 145 LILVKGVEYRMGTITIDDAIGEAADFRTAKLGL--PYP-----GGPAVERAAGNANR 197

QY 62 FDIKPLHHAKNCDFSEFTGLQ-----HYTDKIIMKEKEEGIEKQILSSAADI 110
DB 198 FDPRLVGDARLDFSGIKTAVROAQSIGPVTQDI-----ADV 239
QY 111 AATVQHTNACHLVKRTHRAILFCQKQRLDLPQNNAVLVASGVASNFYIRRALIELTNATQ 170
DB 240 CASFORAISRTRLDRVGRGLKRR-ADFASVDPALVAVAGVANQTLRTTLQSLCDEHG 298
QY 171 CTLLCPPRLCTDNGIMIAWNGIERLARGILHDIGIREPC--PLGVDISKENG 226
DB 299 FRFIAPPLQLCTDNAMIMAGARLAAGL-----PADGLDAAPRSRPLDSEAKALIG 352

Search completed: November 10, 2005, 16:52:03
Job time : 58.9315 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2005, 15:51:52 ; Search time 19.2239 Seconds
(without alignments)
928.069 Million cell updates/sec

Title: US-10-649-273-2_COPY_176_414
Perfect score: 1240
Sequence: 1 LIALVQGVSDPFLIKGSLDI.....DISKVEGASIKVPOLKMEI 239

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: Issued_Patents_AA:*
2: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1240	100.0	267	4	US-10-067-443-22 Sequence 22, Appl
2	1240	100.0	414	4	US-10-067-443-2 Sequence 2, Appl
3	1213.5	97.9	439	4	US-10-067-443-19 Sequence 19, Appl
4	326	26.3	463	4	US-10-067-443-3 Sequence 3, Appl
5	247	19.9	350	4	US-09-540-216-2726 Sequence 2726, Ap
6	231	18.6	401	4	US-09-252-991A-17372 Sequence 17372, A
7	212	17.1	357	4	US-09-543-681A-6513 Sequence 6513, Ap
8	209.5	16.9	342	1	US-08-087-797-3 Sequence 3, Appl
9	205	16.5	431	4	US-09-489-039A-9221 Sequence 9221, Ap
10	201.5	16.2	421	4	US-10-067-443-4 Sequence 4, Appl
11	201.5	16.2	421	1	US-10-067-443-28 Sequence 28, Appl
12	200	16.1	325	1	US-08-087-797-2 Sequence 2, Appl
13	197	15.9	363	4	US-09-107-532A-6609 Sequence 6609, Ap
14	193.5	15.6	327	4	US-10-067-443-5 Sequence 5, Appl
15	190	15.3	336	3	US-08-987-121A-4 Sequence 4, Appl
16	189	15.2	273	4	US-09-710-279-728 Sequence 728, App
17	187	15.1	344	4	US-09-602-777A-148 Sequence 148, App
18	186	15.0	335	3	US-08-961-083-52 Sequence 52, Appl
19	186	15.0	335	4	US-09-536-784-52 Sequence 52, Appl
20	186	15.0	368	3	US-09-134-001C-3909 Sequence 3909, Ap
21	184	14.8	336	3	US-09-066-512-2 Sequence 2, Appl
22	182	14.7	336	4	US-09-583-110-4857 Sequence 4857, Ap
23	181.5	14.6	143	4	US-09-328-352-4387 Sequence 4387, Ap
24	177	14.3	336	4	US-09-107-453-4221 Sequence 4221, Ap
25	174.5	14.1	360	4	US-09-198-452A-213 Sequence 213, App
26	174.5	14.1	360	4	US-09-438-185A-196 Sequence 196, App
27	174	14.0	341	3	US-09-149-624-2 Sequence 2, Appl

28	173	14.0	366	4	US-09-134-000C-4956 Sequence 4956, Ap
29	148.5	12.0	340	4	US-10-067-443-6 Sequence 6, Appl
30	117	9.4	292	4	US-09-724-623-81 Sequence 81, Appl
31	94.5	7.6	480	4	US-09-583-110-5050 Sequence 5050, Ap
32	94.5	7.6	481	4	US-09-107-433-3197 Sequence 3197, Ap
33	89.5	7.2	214	4	US-09-328-352-4609 Sequence 4609, Ap
34	83	6.7	42	4	US-09-902-552-13841 Sequence 13841, A
35	81.5	6.6	1087	4	US-08-264-002-5 Sequence 5, Appl
36	81	6.5	1072	4	US-09-902-540-15572 Sequence 15572, A
37	79	6.4	445	3	US-08-083-945C-2 Sequence 2, Appl
38	79	6.4	445	3	US-08-083-945C-7 Sequence 7, Appl
39	79	6.4	578	4	US-09-252-991A-13318 Sequence 31318, A
40	79	6.4	1137	4	US-09-538-092-968 Sequence 968, App
41	78	6.3	334	1	US-08-530-950-4 Sequence 4, Appl
42	78	6.3	334	3	US-08-576-240-2 Sequence 2, Appl
43	78	6.3	334	3	US-08-888-429A-4 Sequence 4, Appl
44	78	6.3	334	3	US-09-149-879-4 Sequence 4, Appl
45	78	6.3	334	4	US-09-057-009-4 Sequence 4, Appl

ALIGNMENTS

```
RESULT 1
US-10-067-443-22
; Sequence 22, Application US/10067443
; Patent No. 6642041
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED I
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
; LENGTH: 267
; TYPE: PRT
; ORGANISM: homo sapiens
; US-10-067-443-22

Query Match      100.0%; Score 1240; DB 4; Length 267;
Best Local Similarity 100.0%; Pred. No. 5.6e-145;
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 LIALVQGVSDPFLIKGSLDIAPGDMIDKVARRLSLIKPECGTMSGGKAIIEHLAKGNRF 60
      |||
DB      29 LIALVQGVSDPFLIKGSLDIAPGDMIDKVARRLSLIKPECGTMSGGKAIIEHLAKGNRF 88

QY      61 HFDIKPPLHAKNCDFSPFGLOHVTDKIIMKKEKEGIRKGOILSSAADIATVQHTMAC 120
      |||
DB      89 HFDIKPPLHAKNCDFSPFGLOHVTDKIIMKKEKEGIRKGOILSSAADIATVQHTMAC 148

QY      121 HLVKETTRAILFCCKORDLLPQNNAVLVASGVASNYFIRRALEILTNAQTCTLLCPPRL 180
      |||
DB      149 HLVKETTRAILFCCKORDLLPQNNAVLVASGVASNYFIRRALEILTNAQTCTLLCPPRL 208

QY      181 CTDNGIMTAMNGIEHLRAGLGLIHDIIEGIRYEPKCPGLVDISKVEGASIKVPOLKMEI 239
      |||
DB      209 CTDNGIMTAMNGIEHLRAGLGLIHDIIEGIRYEPKCPGLVDISKVEGASIKVPOLKMEI 267

RESULT 2
US-10-067-443-2
; Sequence 2, Application US/10067443
; Patent No. 6642041
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
```


SEQ ID NO 2726
LENGTH: 350
TYPE: PRT
ORGANISM: M. catarrhalis
US-09-540-236-2726

Query Match 19.9%; Score 247; DB 4; Length 350;
Best Local Similarity 31.5%; Pred. No. 1,1e-21;
Matches 70; Conservative 34; Mismatches 80; Indels 38; Gaps 6;

QY 1 LLAIVGVSDFLILGKSLDIAPGDMLDKVARRLSLIKHPECSMTSGGKAIENHAKGNRF 60
DB 149 MLVRADGVGYOILDESIDDAVGECPDKTAKMLK-PIF-----GGPIETKLAKGNPH 201
QY 61 HFIDKPLHAKNCDPFTGLQHTVDKIIIMKEKEGIEKGQILSSAADIATVQHTMAC 120
DB 202 AYELPRPMQH-KGIDFSPGSKMTAIHNLIKOTPMQSDP-----ATRADIASFEVAVVD 255
QY 121 HLVKTRHAILFCRKORDLPONNAVIVASGVASNFYIRALEILTNATQCTLLCPPRL 180
DB 256 TLVKCTALQMTGIRQ-----LVVAGVSAQTLEKRLTETLRQIDASVYVAPTEL 307
QY 181 CTDNGIMIAMNGIERLRAG-----LGILHD 205
DB 308 CTDNGAMIAVAGFCRLSRQSDDLAVRCIPRWDMTMLGIEYD 349

RESULT 6

US-09-252-991A-17372
Sequence 17372, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
PRIOR FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 17372
LENGTH: 401
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17372

Query Match 18.6%; Score 211; DB 4; Length 401;
Best Local Similarity 33.8%; Pred. No. 1.3e-19;
Matches 73; Conservative 34; Mismatches 87; Indels 22; Gaps 8;

QY 2 LALVQGVSDFLILGKSLDIAPGDMLDKVARRLSLIKHPECSMTSGGKAIENHAKGNRF 61
DB 202 LVRVDGIGRYGLDESVDADAAGEAFDKTAKLIGL-GYP-----GGPIETKLAKGNPH 254
QY 62 FDIKPLHAKNCDPFTGLQHTVDKIIIMKEKEGIEKGQILSSAADIATVQHTMAC 121
DB 255 FVPRPMTDRPGDLFSPGSKMTAIHNLIKOTPMQSDP-----ATRADIASFEVAVVD 310
QY 122 LVKTRHAILFCRKORDLPONNAVIVASGVASNFYIRALEILTNATQCTLLCPPRL 181
DB 311 LLIICRRL--KOTGL--K--LVVAGVSAQTLEKRLTETLRQIDASVYVAPTEL 362
QY 182 TDNGIMIAMNGIERLRAGLGIILHDIEGIRYPCPL 217
DB 363 TDNGAMIAVAGFCRLSRQSDDLAVRCIPRWDMTMLGIEYD 395

RESULT 7

US-09-543-681A-6513
Sequence 6513, Application US/09543681A

Patent No. 6605709
GENERAL INFORMATION:
APPLICANT: GARY BRETON
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
FILE REFERENCE: 2709.1002-001
CURRENT APPLICATION NUMBER: US/09/543,681A
PRIOR FILING DATE: 2000-04-05
PRIOR APPLICATION NUMBER: US 60/128,706
PRIOR FILING DATE: 1999-04-09
NUMBER OF SEQ ID NOS: 8344
SEQ ID NO 6513
LENGTH: 357
TYPE: PRT
ORGANISM: Proteus mirabilis
US-09-543-681A-6513

Query Match 17.1%; Score 212; DB 4; Length 357;
Best Local Similarity 30.3%; Pred. No. 2.4e-17;
Matches 66; Conservative 32; Mismatches 90; Indels 30; Gaps 7;

QY 2 LALVQGVSDFLILGKSLDIAPGDMLDKVARRLSLIKHPECSMTSGGKAIENHAKGNRF 61
DB 159 LLSVIGIGRYGLDESIDDAAGEAFDKTAKLIGL-DYP-----GGVLSKMAQQGVEGR 211
QY 62 FDIKPLHAKNCDPFTGLQHTVDKIIIMKEKEGIEKGQILSSAADIATVQHTMAC 121
DB 212 FVPRPMTDRPGDLFSPGSKMTAIHNLIKOTPMQSDP-----ATRADIASFEVAVVD 263
QY 122 LVKTRHAILFCRKORDLPONNAVIVASGVASNFYIRALEILTNATQCTLLCPPRL 180
DB 264 LAKCRRA-----LEQTFKRLVMAGVASANTLRKAMIMIEQLGGEVYVAPTEL 314
QY 181 CTDNGIMIAMNGIERLRAGLGIILHDIEGIRYPCPL 217
DB 315 CTDNGAMIAVAGFCRLSRQSDDLAVRCIPRWDMTMLGIEYD 348

RESULT 8

US-08-087-797-3
Sequence 3, Application US/08087797
Patent No. 5543312
GENERAL INFORMATION:

APPLICANT: Mellors, Alan
APPLICANT: Io, Reggie Y.C.
TITLE OF INVENTION: Pasteurella Haemolytica
TITLE OF INVENTION: Glycoprotease
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSES: Bell, Selzer, Park & Gibson, P.A.
STREET: 1211 East Morehead Street,
CITY: Charlotte
STATE: No. 5543312 Carolina
COUNTRY: United States
ZIP: 28234

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/087,797
FILING DATE: 14-JUL-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Layton, Jr., Samuel G.
REGISTRATION NUMBER: 22807
REFERENCE/DOCKET NUMBER: 3374-80
TELECOMMUNICATION INFORMATION:
TELEPHONE: 704 377 1561
TELEFAX: 704 334 2014

; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 342 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-087-797-3

Query Match 16.9%; Score 209.5; DB 1; Length 342;
Best Local Similarity 30.9%; Pred. No. 4.7e-17;
Matches 67; Conservative 30; Mismatches 97; Indels 23; Gaps 7;

QY 2 LALVGVSDPFLILGSLDIPAGMDLVKVARRLSLIKHEKSGTMSGKAIIEHLAKOGRH 61
DB 142 LLSVIGIQVEYELGSEIDDAAGEAPDKTKLGL-DYP-----GGPLSKRAAAGSTACR 194
QY 62 FDIKPLHAAKNCDFSFTGLQHTVDKIIIMKEKEGIEKGQILSSAADIATVQHTMAC 121
DB 195 FVFPFRPMTDRPGLDPSFGSLKTFPANTIRDNXXXXGXTHDDQ---TRADIARAFEDAVVD 251
QY 122 LVKTRTHAILFCQKQDILPÖNN-AVLVAGSVASNFYIRALEILTNAQTCLLCPPL 180
DB 252 LMIKCKRA-----LDQTGFKRLVMAAGVSANRTIRAKLAEWMKKRGGEVFYARPEF 302
QY 181 CTDNGIMTAMNGIERLRAGLGIHDIIEGIRYEPKCP 217
DB 303 CTDNGAMTAYAGWVRLOGA---GATADL-GVSVRPRMPL 336

RESULT 9

US-09-489-9221
; Sequence 9221, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 9221
; LENGTH: 343
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
; US-09-489-9221

Query Match 16.5%; Score 205; DB 4; Length 343;
Best Local Similarity 30.3%; Pred. No. 1.7e-16;
Matches 66; Conservative 29; Mismatches 93; Indels 30; Gaps 7;

QY 2 LALVGVSDPFLILGSLDIPAGMDLVKVARRLSLIKHEKSGTMSGKAIIEHLAKOGRH 61
DB 148 LLSVIGIQVEYELGSEIDDAAGEAPDKTKLGL-DYP-----GGPLSKRAAAGSTACR 200
QY 62 FDIKPLHAAKNCDFSFTGLQHTVDKIIIMKEKEGIEKGQILSSAADIATVQHTMAC 121
DB 201 FVFPFRPMTDRPGLDPSFGSLKTFPANTIRSDGDE-----QTRADIARAFEDAVVD 252
QY 122 LVKTRTHAILFCQKQDILPÖNN-AVLVAGSVASNFYIRALEILTNAQTCLLCPPL 180
DB 253 LMIKCKRA-----LEQTGFKRLVMAAGVSANRTIRAKLAEWMKKRGGEVFYARPEF 303
QY 181 CTDNGIMTAMNGIERLRAGLGIHDIIEGIRYEPKCP 217
DB 304 CTDNGAMTAYAGWVRLOGA---KAEIGVTVPRMPL 337

RESULT 10

US-10-067-443-4

; Sequence 4, Application US/10067443
; Patent No. 6642041
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 421
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
; US-10-067-443-4

Query Match 16.2%; Score 201.5; DB 4; Length 421;
Best Local Similarity 27.5%; Pred. No. 6.3e-16;
Matches 69; Conservative 48; Mismatches 99; Indels 35; Gaps 7;

QY 1 LALVGVSDPFLILGSLDIPAGMDLVKVARRLSLIKHEKSGTMSGKAIIEHLAKOGRH 58
DB 163 LLSVADVEKFKLXGVSQSGRPECIDKVARQGLD-GSEFSDIHGAANVELASRASAD 221
QY 59 -RHFIDIKPLHAAKNCDFSFTGLQHTVDKIIIMKEKEGIEKGQILSSAADIATVQHT 117
DB 222 GHLRYDIFLPNVPKAMNPDQIKGSLYNTLERLRKNSSETID-----IPDFCASLQNT 274
QY 118 MACHLVKTRTHAILFCQKQDILPÖNNAVLVAGSVASNFYIRALEILTNAQTCLLCP 177
DB 275 VARHISSKKHIFPESLSEQEKLPKQ---LVIGGVANQYIFGAISKLSAHHVTTIKVL 331
QY 178 PLCTDNGIMTAMNGIERLRAGLGIHDIIEGIRYEP-----KCPAGVDISKEVG 226
DB 332 LSLCTDNGIMTAMNGIERLRAGLGIHDIIEGIRYEP-----KCPAGVDISKEVG 226
QY 227 EASIKVPQDKM 237
DB 385 ---IDPRRKL 392

RESULT 11

US-10-067-443-28
; Sequence 28, Application US/10067443
; Patent No. 6642041
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 28
; LENGTH: 421
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
; US-10-067-443-28

Query Match 16.2%; Score 201.5; DB 4; Length 421;
Best Local Similarity 27.5%; Pred. No. 6.3e-16;
Matches 69; Conservative 48; Mismatches 99; Indels 35; Gaps 7;

QY 1 LALVGVSDPFLIGSLDIPGMDLVKVARLSLIKPECSMTSGGKAIHILAKQGN-- 58
DB 163 LLSVABEVEKFKLYGGVSSGPGCIDKVAROLDL-GSEFPGIVGAVENTLASRASAD 221
QY 59 -RHFHDKPPLHNAKNCDFSTGLQHTVDKIIMKEKEGIEKGQILSSADIAATVQHT 117
DB 222 GHLRYPFLPNVPANANFDQIKSYLNLRLRKNSTSID-----IPDFCSLDQNT 274
QY 118 MACHLVKTRRALIFCKORDLLPONNAVLVASGGVASFYIRALEILTNAOCTLLCP 177
DB 275 VARIHISKHLIPFESLSEQELPKQ---LVIGGVAANOYIFGATSKLSAHNVTITIKL 331
QY 178 PRLCTDNGIMIANNGIERLRAGLGLHDIGIRYEP-----KCPGLVDISKEVG 226
DB 332 LSLCTDPAEMIAVSGL-----LMLVNRSEAIMWRPNIDPITYAHARSDIGTDASSEI- 384
QY 227 EASIKVPQLKM 237
DB 385 ---IDTPRRKL 392

RESULT 12

US-08-087-797-2
Sequence 2, Application US/08087797
Patent No. 5543312
GENERAL INFORMATION:
APPLICANT: Mellors, Alan
APPLICANT: lo, Reggie Y.C.
APPLICANT: Abdullah, Khalid M.
TITLE OF INVENTION: Pasteurella Haemolytica
TITLE OF INVENTION: Glycoprotease
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Bell, Seltzer, Park & Gibson, P.A.
STREET: 1211 East Morehead Street,
CITY: Charlotte
STATE: No. 5543312ch Carolina
COUNTRY: United States
ZIP: 28234
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/087,797
FILING DATE: 14-JUL-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Layton, Jr., Samuel G.
REGISTRATION NUMBER: 22807
REFERENCE/DOCKET NUMBER: 3374-80
TELECOMMUNICATION INFORMATION:
TELEPHONE: 704 377 1561
TELEFAX: 704 334 2014
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 325 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-087-797-2

Query Match 16.1%; Score 200; DB 1; Length 325;
Best Local Similarity 31.6%; Pred. No. 6.5e-16;
Matches 62; Conservative 25; Mismatches 91; Indels 18; Gaps 5;

QY 2 LALVGVSDPFLIGSLDIPGMDLVKVARLSLIKPECSMTSGGKAIHILAKQGNRF 61
DB 142 LVVVDGQGYELIGESTIDDAAGEAFDKTKLGL-----AGVAMSKLASGTPNR 194
QY 62 FDIKPLHNAKNCDFSTGLQHTVDKIIMKEKEGIEKGQILSSADIAATVQHTMA 121

DB 195 FRPFRMTDRPGLDFSGIKTPAANTIRANINENELDEQ---TKCDIAHAFQOAV--- 248
QY 122 LVKTRRALIFCKORDLLPONNAVLVASGGVASFYIRALEILTNAOCTLLCPRLC 181
DB 249 ---VDTIKCK-RALBQTKYKLVAGGVASANKQRLADLAEKMKLKGVEVYPRQFC 303
QY 182 TDNGIMIANNGIERLR 197
DB 304 TDNGAMIAVTGLRLK 319

RESULT 13

US-09-107-532A-6609
Sequence 6609, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02154
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 6609:
SEQUENCE CHARACTERISTICS:
LENGTH: 363 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (8) LOCATION 1...363
SEQUENCE DESCRIPTION: SEQ ID NO: 6609:
US-09-107-532A-6609

Query Match 15.9%; Score 197; DB 4; Length 363;
Best Local Similarity 32.6%; Pred. No. 1.8e-15;
Matches 63; Conservative 34; Mismatches 64; Indels 32; Gaps 9;

QY 2 LALVGVSDPFLIGSLDIPGMDLVKVARLSLIKPECSMTSGGKAIHILAKQGNRF 60
DB 169 LVVVDGQGYELIGESTIDDAAGEAFDKTKLGL-----SYPSKREIDQLAHQCKDNY 221
QY 61 HFDIKPLHNAKNCDFSTGLQHTVDKIIMK-KEKEGIEKGQILSSADIAATVQHTMA 119

Db 222 HF--PRAMIEDNDYDFSGLSKSAFINLVHNAQORGEIDLKN-----DLAASFOASVI 272
Qy 120 CHLVKTRTRAILFCQKQDLPLPQNNNAV--LVASGVASNFYIRALEILITNA--TOCTLLC 175
Db 273 DVLTITKTLRA---C-----QNTYVVKQLVVAAGVAAQAQLEGLQALASATLPEVELVI 322
Qy 176 PPPRLCTDNGIMI 188
Db 323 PPLRLCGDNAMI 335

RESULT 14
US-10-067-443-5
; Sequence 5, Application US/10067443
; Patent No. 6642041
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE HIGHLY EXPRESSED IN
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
; LENGTH: 327
; TYPE: PRT
; ORGANISM: Thermotoga maritima
US-10-067-443-5

Query Match 15.6%; Score 193.5; DB 4; Length 327;
Best Local Similarity 28.9%; Pred. No. 4.2e-15;
Matches 55; Conservative 38; Mismatches 68; Indels 29; Gaps 6;

Qy 13 LIGKSLDIAPGMDLKVARRSLIKHPECSTWGGKATIEHLAKQGNRFHDIKPLHNK 72
Db 151 VLGETLDLSAGEAFKVARLGL-GYP-----GGPVLDVAKKGDPEKYSFPPRLDD 203
Qy 73 NCDPSFTGLQHTDKIIMKKEKEGIEKQILSSADIATVQHTMACHLVKTRTRAILF 132
Db 204 SYNSFPAIK---TSVLYFLQREKGYK-----VEDVASQKAVVDLVEKTRP----- 249
Qy 133 CKQKDLPLPQNNAV--LVASGVASNFYIRALEILITNAQCTLLCPPRLCTDNGIMIAW 190
Db 250 -----LARNIGIRKIAFVGVAANSMLREVRKRAERNVYVFPPLELCTDNALMVAK 303
Qy 191 NGIERLRAGL 200
Db 304 AGYEKAKRGM 313

RESULT 15
US-08-987-121A-4
; Sequence 4, Application US/08987121A
; GENERAL INFORMATION:
; APPLICANT: Hoskins, Jo Ann
; APPLICANT: Tang, Joseph Chiou-Chung
; APPLICANT: Treadway, Patli Jean
; TITLE OF INVENTION: Streptococcus Pneumoniae Gene Sequence
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: U.S.
; ZIP: 46285
; COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/987,121A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Webster, Thomas D.
; REGISTRATION NUMBER: 39,872
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-3334
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 336 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-987-121A-4

Query Match 15.3%; Score 190; DB 3; Length 336;
Best Local Similarity 31.1%; Pred. No. 1.2e-14;
Matches 61; Conservative 32; Mismatches 73; Indels 30; Gaps 8;

Qy 2 LALVGVSDPFLILGKSLDIAPGMDLKVARRSLIKHPECSTWGGKATIEHLAKQGNRFH 61
Db 144 LVYVSEAGDKYIVGETRDAVGEAYDKVGRWGL-----TYPAREIDELAHQCHDI- 195
Qy 62 FDIKPEPLHAKKCDPSFTGLQHTVDKIIMKKEKEGIEKQILSSADIATVQHTMACH 121
Db 196 YDFPRAMIKEDNLEFSFGLSKSAFINLVHNAE-----QKGESIST-EDLCASFQAAVMDI 249
Qy 122 LVKTRTRAILFCQKQDLPLPQNNNAV--LVASGVASNFYIRALEILITNAQCTLLCPPRLC 181
Db 250 LMAKTKKAL-----EKYPVK--TLVAVAGVANKGLRRL--ATBITDVNVIIPPLRLC 299
Qy 182 TDNGIMIA-----NN 191
Db 300 GDNAGMIAVAVSEWN 315

Search completed: November 10, 2005, 16:54:35
Job time : 20.2239 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2005, 16:43:58 ; Search time 63.1272 Seconds

(without alignments)
1584.102 Million cell updates/sec

Title: US-10-649-273-2_COPY_176_414

Perfect score: 1240
Sequence: 1 LIALVQGVSDPFLIGKSLDI.....DISKEVGASIKVQLKMEI 239

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1867879 seqs, 418409474 residues

Total number of hits satisfying chosen parameters: 1867879

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09C_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US10E_PUBCOMB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
19: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep:*
20: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep:*
21: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*
22: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1240	100.0	267	14	US-10-067-443-22 Sequence 22, Appl
2	1240	100.0	267	15	US-10-649-273-22 Sequence 22, Appl
3	1240	100.0	267	15	US-10-651-722-22 Sequence 22, Appl
4	1240	100.0	414	14	US-10-067-443-2 Sequence 2, Appl
5	1240	100.0	414	15	US-10-649-273-2 Sequence 2, Appl
6	1240	100.0	414	15	US-10-651-722-2 Sequence 2, Appl
7	1240	100.0	414	17	US-10-480-988-8 Sequence 8, Appl
8	1213.5	97.9	439	15	US-10-651-722-19 Sequence 19, Appl
9	1213.5	97.9	439	15	US-10-649-273-19 Sequence 19, Appl
10	1213.5	97.9	439	15	US-10-651-722-19 Sequence 19, Appl
11	1203	97.0	414	14	US-10-012-140-5 Sequence 5, Appl

12	960	77.4	364	15	US-10-094-749-2039	Sequence 2039, Ap
13	337	27.2	445	15	US-10-424-599-209259	Sequence 209259,
14	332.5	26.8	409	20	US-11-097-143-34191	Sequence 14191, A
15	326	26.3	463	14	US-10-067-443-3	Sequence 3, Appl
16	326	26.3	463	15	US-10-649-273-3	Sequence 3, Appl
17	326	26.3	463	15	US-10-651-722-3	Sequence 3, Appl
18	270	21.8	444	16	US-10-437-963-113732	Sequence 113732,
19	257	20.7	382	15	US-10-282-122A-50858	Sequence 50858, A
20	254	20.5	333	14	US-10-012-140-25	Sequence 25, Appl
21	248.5	20.0	348	15	US-10-282-122A-63156	Sequence 63156, A
22	247	19.9	312	15	US-10-282-122A-44499	Sequence 44499, A
23	241	19.4	343	15	US-10-282-122A-67227	Sequence 67227, A
24	239	19.3	251	14	US-10-081-051-9	Sequence 11798, A
25	231	18.6	341	9	US-09-815-242-11798	Sequence 66200, A
26	231	18.6	341	15	US-10-282-122A-66200	Sequence 52, Appl
27	231	18.6	341	18	US-10-958-216-50	Sequence 50, Appl
28	229	18.5	341	18	US-10-958-216-52	Sequence 24, Appl
29	228.5	18.4	347	14	US-10-012-140-24	Sequence 11043, A
30	224	18.1	342	9	US-09-815-242-11043	Sequence 58204, A
31	224	18.1	342	15	US-10-282-122A-58204	Sequence 10128, A
32	213.5	17.2	347	20	US-11-097-143-10128	Sequence 13780, A
33	213	17.2	337	9	US-09-815-242-13780	Sequence 75485, A
34	213	17.2	337	15	US-10-282-122A-75485	Sequence 68438, A
35	212	17.1	340	15	US-10-282-122A-68438	Sequence 26, Appl
36	211	17.0	41	14	US-10-012-140-26	Sequence 78161, A
37	210	16.9	337	15	US-10-282-122A-78161	Sequence 55404, A
38	208	16.8	335	15	US-10-282-122A-55404	Sequence 67993, A
39	207.5	16.7	341	15	US-10-282-122A-67993	Sequence 111620, A
40	207	16.7	441	16	US-10-437-963-111620	Sequence 10304, A
41	206	16.6	337	9	US-09-815-242-10304	Sequence 56695, A
42	206	16.6	337	15	US-10-282-122A-56695	Sequence 67198, A
43	206	16.6	396	15	US-10-425-114-67198	Sequence 4, Appl
44	201.5	16.2	421	14	US-10-067-443-4	Sequence 28, Appl
45	201.5	16.2	421	14	US-10-067-443-28	

ALIGNMENTS

RESULT 1
US-10-067-443-22
; Sequence 22, Application US/10067443
; Publication No. US20030082782A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED I
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
; LENGTH: 267
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-067-443-22

Query Match 100.0%; Score 1240; DB 14; Length 267;
Best Local Similarity 100.0%; Pred. No. 2.3e-127;
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LIALVQGVSDPFLIGKSLDIAPCDMDKXVARRSLIKHPGCSMGSGKAIENHAKGNRP 60
DB 29 LIALVQGVSDPFLIGKSLDIAPCDMDKXVARRSLIKHPGCSMGSGKAIENHAKGNRP 60
QY 61 HFDIKPPLHAKRCDFSPFGLOHVTDKIIMKCKEKGIEKGQITLSSAADIATVQHTMAC 120
DB 89 HFDIKPPLHAKRCDFSPFGLOHVTDKIIMKCKEKGIEKGQITLSSAADIATVQHTMAC 148

QY 121 HLKRTTRAILFCCKORDLLPQNNAVLVASGVASNFYIRRALEILTNATQCTLLCPPRL 180
DB 149 HLKRTTRAILFCCKORDLLPQNNAVLVASGVASNFYIRRALEILTNATQCTLLCPPRL 208
QY 181 CTDNGIMIAMNGIERLRAGLGIHLDIEGIRYEPKCPGLVDISKVEGASIKVPOLKMEI 239
DB 209 CTDNGIMIAMNGIERLRAGLGIHLDIEGIRYEPKCPGLVDISKVEGASIKVPOLKMEI 267

RESULT 2

US-10-649-273-22
; Sequence 22, Application US/10649273
; Publication No. US20040043407A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1
; FILE REFERENCE: D0073 CNT
; CURRENT APPLICATION NUMBER: US/10/649,273
; CURRENT FILING DATE: 2003-08-27
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 22
; LENGTH: 267
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-649-273-22

Query Match 100.0%; Score 1240; DB 15; Length 267;
Best Local Similarity 100.0%; Pred. No. 2.3e-127;
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLAIVQVSDPFLILGKSLDIAPGMDLKVARRSLIKHPECSTMSGKAI EHLAKGNRF 60
DB 29 LLAIVQVSDPFLILGKSLDIAPGMDLKVARRSLIKHPECSTMSGKAI EHLAKGNRF 88
QY 61 HFDIKPPLHAKNCDPFTGLQHTVDKIIIMKKEKEGIEKGQILSSAADIAATVQHTMAC 120
DB 89 HFDIKPPLHAKNCDPFTGLQHTVDKIIIMKKEKEGIEKGQILSSAADIAATVQHTMAC 148
QY 121 HLKRTTRAILFCCKORDLLPQNNAVLVASGVASNFYIRRALEILTNATQCTLLCPPRL 180
DB 149 HLKRTTRAILFCCKORDLLPQNNAVLVASGVASNFYIRRALEILTNATQCTLLCPPRL 208
QY 181 CTDNGIMIAMNGIERLRAGLGIHLDIEGIRYEPKCPGLVDISKVEGASIKVPOLKMEI 239
DB 209 CTDNGIMIAMNGIERLRAGLGIHLDIEGIRYEPKCPGLVDISKVEGASIKVPOLKMEI 267

RESULT 3

US-10-651-722-22
; Sequence 22, Application US/10651722
; Publication No. US20040048302A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1
; FILE REFERENCE: D0073 DIV
; CURRENT APPLICATION NUMBER: US/10/651,722
; CURRENT FILING DATE: 2003-08-29
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 22
; LENGTH: 267
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-651-722-22

Query Match 100.0%; Score 1240; DB 15; Length 267;
Best Local Similarity 100.0%; Pred. No. 2.3e-127;
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLAIVQVSDPFLILGKSLDIAPGMDLKVARRSLIKHPECSTMSGKAI EHLAKGNRF 60
DB 29 LLAIVQVSDPFLILGKSLDIAPGMDLKVARRSLIKHPECSTMSGKAI EHLAKGNRF 88
QY 61 HFDIKPPLHAKNCDPFTGLQHTVDKIIIMKKEKEGIEKGQILSSAADIAATVQHTMAC 120
DB 89 HFDIKPPLHAKNCDPFTGLQHTVDKIIIMKKEKEGIEKGQILSSAADIAATVQHTMAC 148
QY 121 HLKRTTRAILFCCKORDLLPQNNAVLVASGVASNFYIRRALEILTNATQCTLLCPPRL 180
DB 149 HLKRTTRAILFCCKORDLLPQNNAVLVASGVASNFYIRRALEILTNATQCTLLCPPRL 208
QY 181 CTDNGIMIAMNGIERLRAGLGIHLDIEGIRYEPKCPGLVDISKVEGASIKVPOLKMEI 239
DB 209 CTDNGIMIAMNGIERLRAGLGIHLDIEGIRYEPKCPGLVDISKVEGASIKVPOLKMEI 267

RESULT 4

US-10-067-443-2
; Sequence 2, Application US/10067443
; Publication No. US20030082782A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN
; FILE REFERENCE: SPINAL CORD, MP-1
; CURRENT APPLICATION NUMBER: US/10/067,443
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-067-443-2

Query Match 100.0%; Score 1240; DB 14; Length 414;
Best Local Similarity 100.0%; Pred. No. 4.3e-127;
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLAIVQVSDPFLILGKSLDIAPGMDLKVARRSLIKHPECSTMSGKAI EHLAKGNRF 60
DB 176 LLAIVQVSDPFLILGKSLDIAPGMDLKVARRSLIKHPECSTMSGKAI EHLAKGNRF 235
QY 61 HFDIKPPLHAKNCDPFTGLQHTVDKIIIMKKEKEGIEKGQILSSAADIAATVQHTMAC 120
DB 236 HFDIKPPLHAKNCDPFTGLQHTVDKIIIMKKEKEGIEKGQILSSAADIAATVQHTMAC 295
QY 121 HLKRTTRAILFCCKORDLLPQNNAVLVASGVASNFYIRRALEILTNATQCTLLCPPRL 180
DB 296 HLKRTTRAILFCCKORDLLPQNNAVLVASGVASNFYIRRALEILTNATQCTLLCPPRL 355
QY 181 CTDNGIMIAMNGIERLRAGLGIHLDIEGIRYEPKCPGLVDISKVEGASIKVPOLKMEI 239
DB 356 CTDNGIMIAMNGIERLRAGLGIHLDIEGIRYEPKCPGLVDISKVEGASIKVPOLKMEI 414

RESULT 5

US-10-649-273-2


```
; Sequence 2, Application US/10649273
; Publication No. US20040043407A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1
; FILE REFERENCE: D0073 CNT
; CURRENT APPLICATION NUMBER: US/10/649,273
; CURRENT FILING DATE: 2003-08-27
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-649-273-2

Query Match          100.0%; Score 1240; DB 15; Length 414;
Best Local Similarity 100.0%; Pred. No. 4.3e-127;
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLAIVQGVSDPFLILKSLDIAPGDMLDKVARSLTIKHEPCSTMSGKAIIEHLAKQGNRF 60
    |||||
Db 176 LLAIVQGVSDPFLILKSLDIAPGDMLDKVARSLTIKHEPCSTMSGKAIIEHLAKQGNRF 235
Qy 61 HPDIKPLHAKNCDFSTGLQHTVDKTIIMKKEKEGIEKGQILSSAADIATVQHTMAC 120
Db 236 HPDIKPLHAKNCDFSTGLQHTVDKTIIMKKEKEGIEKGQILSSAADIATVQHTMAC 295
Qy 121 HLKRTTRAILFCQKRDLLPQNNAVLVASGVASNFYIRALIEILTNATQCTLLCPPRL 180
Db 296 HLKRTTRAILFCQKRDLLPQNNAVLVASGVASNFYIRALIEILTNATQCTLLCPPRL 355
Qy 181 CTDNGIMIAMNGIERLRAGLGIHDIIEGIRYKPCPLGVDSKEVGEASIKVPOLKMEI 239
Db 356 CTDNGIMIAMNGIERLRAGLGIHDIIEGIRYKPCPLGVDSKEVGEASIKVPOLKMEI 414

RESULT 6
US-10-651-722-2
; Sequence 2, Application US/10651722
; Publication No. US20040048302A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1
; FILE REFERENCE: D0073 DIV
; CURRENT APPLICATION NUMBER: US/10/651,722
; CURRENT FILING DATE: 2003-08-29
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-651-722-2

Query Match          100.0%; Score 1240; DB 15; Length 414;
Best Local Similarity 100.0%; Pred. No. 4.3e-127;
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLAIVQGVSDPFLILKSLDIAPGDMLDKVARSLTIKHEPCSTMSGKAIIEHLAKQGNRF 60
    |||||
```

```
Db 176 LLAIVQGVSDPFLILKSLDIAPGDMLDKVARSLTIKHEPCSTMSGKAIIEHLAKQGNRF 235
Qy 61 HPDIKPLHAKNCDFSTGLQHTVDKTIIMKKEKEGIEKGQILSSAADIATVQHTMAC 120
Db 236 HPDIKPLHAKNCDFSTGLQHTVDKTIIMKKEKEGIEKGQILSSAADIATVQHTMAC 295
Qy 121 HLKRTTRAILFCQKRDLLPQNNAVLVASGVASNFYIRALIEILTNATQCTLLCPPRL 180
Db 296 HLKRTTRAILFCQKRDLLPQNNAVLVASGVASNFYIRALIEILTNATQCTLLCPPRL 355
Qy 181 CTDNGIMIAMNGIERLRAGLGIHDIIEGIRYKPCPLGVDSKEVGEASIKVPOLKMEI 239
Db 356 CTDNGIMIAMNGIERLRAGLGIHDIIEGIRYKPCPLGVDSKEVGEASIKVPOLKMEI 414

RESULT 7
US-10-480-988-8
; Sequence 8, Application US/10480988
; Publication No. US20050069877A1
; GENERAL INFORMATION:
; APPLICANT: GANDHI, Ameena R.; KABLE, Amy E.;
; APPLICANT: SWARNAKAR, Anita; HAPALIA, April J.A.;
; APPLICANT: TRAN, Bao; DUGGAN, Brendan M.;
; APPLICANT: WARREN, Bridget A.; ISON, Craig H.;
; APPLICANT: HONCHELL, Cynthia D.; NGUYEN, Daniel B.;
; APPLICANT: LU, Dyung Aina M.; LEE, Ernestine A.;
; APPLICANT: YUE, Henry; FORSYTHE, Ian J.;
; APPLICANT: BARROSO, Ines; RAMKUMAR, Jayalaxmi;
; APPLICANT: GRIFFIN, Jennifer A.; LI, Yoon X.;
; APPLICANT: YANG, Junming; THANGAVELOU, Kavitha;
; APPLICANT: GIETZEN, Kimberley J.; DING, Li;
; APPLICANT: BAUGHN, Mariah R.; BOROMSKY, Mark L.;
; APPLICANT: YAO, Monique G.; CHAWLA, Narinder K.;
; APPLICANT: MASON, Patricia M.; GURURAJAN, Rajagopal;
; APPLICANT: LEE, Sally; BECHA, Shanya D.;
; APPLICANT: LEE, Soo Yeun; TRAN, Uyen K.;
; APPLICANT: ELIOTT, Vicki S.; LUO, Wen;
; APPLICANT: SPRAGUE, William W.; TANG, Y. Tom;
; APPLICANT: LU, Yan; ZEBARADIAN, Yeganeh
; TITLE OF INVENTION: PROTEIN MODIFICATION AND MAINTENANCE MOLECULES
; FILE REFERENCE: PF-1040 USN
; CURRENT APPLICATION NUMBER: US/10/480,988
; CURRENT FILING DATE: 2003-12-16
; PRIOR APPLICATION NUMBER: PCT/US02/19360
; PRIOR FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: US 60/300,508
; PRIOR FILING DATE: 2001-06-22
; PRIOR APPLICATION NUMBER: US 60/303,445
; PRIOR FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US 60/305,405
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/311,442
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: US 60/314,821
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/315,992
; PRIOR FILING DATE: 2001-08-29
; PRIOR APPLICATION NUMBER: US 60/378,205
; PRIOR FILING DATE: 2002-05-03
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: PERL Program
; SEQ ID NO 8
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No: 7632424CD1
US-10-480-988-8

Query Match          100.0%; Score 1240; DB 17; Length 414;
Best Local Similarity 100.0%; Pred. No. 4.3e-127;
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

QY      1 LALVGVSDPFLILGKSLDIAPGDMLDKVARRLSLIKHPECS TMSGGKAI EHLAKQGNRF 60
      176 LLAIVGVSDPFLILGKSLDIAPGDMLDKVARRLSLIKHPECS TMSGGKAI EHLAKQGNRF 235
DB      61 HFDIKPPLHAAKNCDFSFTGLQHTVDKTIIMKEKEEGIEKGOILSSAADIAATVQHTMAC 120
QY      236 HFDIKPPLHAAKNCDFSFTGLQHTVDKTIIMKEKEEGIEKGOILSSAADIAATVQHTMAC 295
DB      121 HLVRTRRALIFCQORDLLPQNNAVLVASGVASNPYIRALEITLTNATQCTLLCPPRL 180
QY      296 HLVRTRRALIFCQORDLLPQNNAVLVASGVASNPYIRALEITLTNATQCTLLCPPRL 355
DB      181 CTDNGIMIANMNGIERLRAGLGIHIDIEGIRYEPKCPILGVDSKEVGEASIKVPOLKMEI 239
QY      356 CTDNGIMIANMNGIERLRAGLGIHIDIEGIRYEPKCPILGVDSKEVGEASIKVPOLKMEI 414

```

RESULT 8

```

US-10-067-443-19
; Sequence 19, Application US/10067443
; Publication No. US20030082782A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE HIGHLY EXPRESSED IN
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 19
; LENGTH: 439
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-067-443-19

```

```

Query Match      97.9%; Score 1213.5; DB 14; Length 439;
Best Local Similarity 90.2%; Pred. No. 3.9e-124;
Matches 238; Conservative 0; Mismatches 1; Indels 25; Gaps 1;

```

```

QY      1 LLAIVGVSDPFLILGKSLDIAPGDMLDKVARRLSLIKHPECS TMSGGKAI EHLAKQGNRF 60
      176 LLAIVGVSDPFLILGKSLDIAPGDMLDKVARRLSLIKHPECS TMSGGKAI EHLAKQGNRF 235
DB      61 HFDIKPPLHAAKNCDFSFTGLQHTVDKTIIMKEKEEGI----- 98
QY      236 HFDIKPPLHAAKNCDFSFTGLQHTVDKTIIMKEKEEGIFLISKVEQINIPGLCLKIAAHF 295
DB      99 ---EKGQILSSAADIAATVQHTMACHLVTRTRRALIFCQORDLLPQNNAVLVASGVASN 155
QY      296 CRVEKGOILSSAADIAATVQHTMACHLVTRTRRALIFCQORDLLPQNNAVLVASGVASN 355
DB      156 FYIRRALEITLTNATQCTLLCPPRLCTDNGIMIANMNGIERLRAGLGIHIDIEGIRYEPKC 215
QY      356 FYIRRALEITLTNATQCTLLCPPRLCTDNGIMIANMNGIERLRAGLGIHIDIEGIRYEPKC 415
DB      216 PLGVDSKEVGEASIKVPOLKMEI 239
QY      416 PLGVDSKEVGEASIKVPOLKMEI 439

```

RESULT 9

```

US-10-649-273-19
; Sequence 19, Application US/10649273
; Publication No. US20040043407A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1

```

```

; FILE REFERENCE: D0073 CNT
; CURRENT APPLICATION NUMBER: US/10/649,273
; CURRENT FILING DATE: 2003-08-27
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 19
; LENGTH: 439
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-649-273-19

```

```

Query Match      97.9%; Score 1213.5; DB 15; Length 439;
Best Local Similarity 90.2%; Pred. No. 3.9e-124;
Matches 238; Conservative 0; Mismatches 1; Indels 25; Gaps 1;

```

```

QY      1 LLAIVGVSDPFLILGKSLDIAPGDMLDKVARRLSLIKHPECS TMSGGKAI EHLAKQGNRF 60
      176 LLAIVGVSDPFLILGKSLDIAPGDMLDKVARRLSLIKHPECS TMSGGKAI EHLAKQGNRF 235
DB      61 HFDIKPPLHAAKNCDFSFTGLQHTVDKTIIMKEKEEGI----- 98
QY      236 HFDIKPPLHAAKNCDFSFTGLQHTVDKTIIMKEKEEGIFLISKVEQINIPGLCLKIAAHF 295
DB      99 ---EKGQILSSAADIAATVQHTMACHLVTRTRRALIFCQORDLLPQNNAVLVASGVASN 155
QY      296 CRVEKGOILSSAADIAATVQHTMACHLVTRTRRALIFCQORDLLPQNNAVLVASGVASN 355
DB      156 FYIRRALEITLTNATQCTLLCPPRLCTDNGIMIANMNGIERLRAGLGIHIDIEGIRYEPKC 215
QY      356 FYIRRALEITLTNATQCTLLCPPRLCTDNGIMIANMNGIERLRAGLGIHIDIEGIRYEPKC 415
DB      216 PLGVDSKEVGEASIKVPOLKMEI 239
QY      416 PLGVDSKEVGEASIKVPOLKMEI 439

```

RESULT 10

```

US-10-651-722-19
; Sequence 19, Application US/10651722
; Publication No. US20040048302A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1
; FILE REFERENCE: D0073 DIV
; CURRENT APPLICATION NUMBER: US/10/651,722
; CURRENT FILING DATE: 2003-08-29
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 19
; LENGTH: 439
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-651-722-19

```

```

Query Match      97.9%; Score 1213.5; DB 15; Length 439;
Best Local Similarity 90.2%; Pred. No. 3.9e-124;
Matches 238; Conservative 0; Mismatches 1; Indels 25; Gaps 1;

```

```

QY      1 LLAIVGVSDPFLILGKSLDIAPGDMLDKVARRLSLIKHPECS TMSGGKAI EHLAKQGNRF 60
      176 LLAIVGVSDPFLILGKSLDIAPGDMLDKVARRLSLIKHPECS TMSGGKAI EHLAKQGNRF 235

```

```
QY 61 HFDIKPPLHAKNCDFSTGLQHTVDKIIMKKEKEGKI----- 98
D 236 HFDIKPPLHAKNCDFSTGLQHTVDKIIMKKEKEGKIPLISKEQINI PGLCKIAHF 295
QY 99 ---EKGOILSSADIAATVOHTMACILVYKTRHAILFCRKORDLLPNNNAVVASGVASN 155
D 296 CRYEKQILSSADIAATVOHTMACILVYKTRHAILFCRKORDLLPNNNAVVASGVASN 355
QY 156 FYIRRALBELTNATQCTLLCPPRLCTDNGIMIANWGIERLRLAGLILHDIEGIRYEPKC 215
D 356 FYIRRALBELTNATQCTLLCPPRLCTDNGIMIANWGIERLRLAGLILHDIEGIRYEPKC 415
QY 216 PLGVDISKEVGEASIKVPOLKMEI 239
D 416 PLGVDISKEVGEASIKVPOLKMEI 439

RESULT 11
US-10-012-140-5
; Sequence 5, Application US/10012140
; Publication No. US20030009017A1
; GENERAL INFORMATION:
; APPLICANT: Leibyl, Kevin R.
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Glucksmann, Maria A.
; TITLE OF INVENTION: 38650, 28472, 5495, 65507, 81588, AND
; TITLE OF INVENTION: 14354 METHODS AND COMPOSITIONS OF HUMAN PROTEINS AND USES
; TITLE OF INVENTION: THERSOP
; FILE REFERENCE: 381552004500
; CURRENT APPLICATION NUMBER: US/10/012,140
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: 60/246,768
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/246,772
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/249,185
; PRIOR FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-012-140-5

Query Match 97.0%; Score 1203; DB 14; Length 414;
Best Local Similarity 97.1%; Pred. No. 5,1e-123;
Matches 232; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 LLAIVGVSDPFLILGSLDIPGMDLKVARRSLIKHPECSTMSGKAI EHLAKQGNRF 60
D 176 LLAIVGVSDPFLILGSLDIPGMDLKVARRSLIKHPECSTMSGKAI EHLAKQGNRF 235
QY 61 HFDIKPPLHAKNCDFSTGLQHTVDKIIMKKEKEGKIPLISADIAATVOHTMAC 120
D 236 HFDIKPPLHAKNCDFSTGLQHTVDKIIMKKEKEGKIPLISADIAATVOHTMAC 295
QY 121 HLVKTRHAILFCRKORDLLPNNNAVVASGVASNYIRRALBELTNATQCTLLCPPRL 180
D 296 HLVKTRHAILFCRKORDLLPNNNAVVASGVASNYIRRALBELTNATQCTLLCPPRL 355
QY 181 CTDNGIMIANWGIERLRLAGLILHDIEGIRYEPKCPLGVDISKEVGEASIKVPOLKMEI 239
D 356 CTDNGIMIANWGIERLRLAGLILHDIEGIRYEPKCPLGVDISKEVGEASIKVPOLKMEI 414

RESULT 12
US-10-094-749-2039
; Sequence 2039, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
```

```
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHITO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOKUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIRO
; TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; PRIOR FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2039
; LENGTH: 364
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-749-2039

Query Match 77.4%; Score 960; DB 15; Length 364;
Best Local Similarity 98.4%; Pred. No. 2,2e-96;
Matches 186; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 LLAIVGVSDPFLILGSLDIPGMDLKVARRSLIKHPECSTMSGKAI EHLAKQGNRF 60
D 176 LLAIVGVSDPFLILGSLDIPGMDLKVARRSLIKHPECSTMSGKAI EHLAKQGNRF 235
QY 61 HFDIKPPLHAKNCDFSTGLQHTVDKIIMKKEKEGKIPLISADIAATVOHTMAC 120
D 236 HFDIKPPLHAKNCDFSTGLQHTVDKIIMKKEKEGKIPLISADIAATVOHTMAC 295
QY 121 HLVKTRHAILFCRKORDLLPNNNAVVASGVASNYIRRALBELTNATQCTLLCPPRL 180
D 296 HLVKTRHAILFCRKORDLLPNNNAVVASGVASNYIRRALBELTNATQCTLLCPPRL 355
QY 181 CTDNGIMIA 189
D 356 CTDNGIMIA 364

RESULT 13
US-10-424-598-209259
; Sequence 209259, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; PRIOR FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 209259
; LENGTH: 445
; TYPE: PRT
; ORGANISM: Glycine max
```

```
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT3847_3098C.1.pep
US-10-424-599-209259

Query Match      27.2%; Score 337; DB 15; Length 445;
Best Local Similarity 38.5%; Pred. No. 8.9e-28;
Matches 85; Conservative 29; Mismatches 83; Indels 24; Gaps 5;

QY 1 LVALVGVSDPFLILGKSLDIAPGMDLVKVARSLIKHPECSTMSGKAIETHAKQGNRF 60
DB 194 LVLARLDIGQYIQLGTTTDDAIGEAIDYTKAKMGLDLR-----RSGGPAIEKLAEGNME 248
QY 61 HFDIKPPLHAKNCDSEFTGLQ----HYTDKIIMKEKEGIEKQILSSAADIATVQ 115
DB 249 SVKFSIPMKHKDCNFSYAGLKTQVRLAIESKKIDAKIPISSANGDRL-SRADIASPQ 307
QY 116 HTMACHLVKTRRAILFCQKQDLPQNNNAVLSGGVANSFYIRALBITLTAQCTLLIC 175
DB 308 RIAVLHLEERGERALQWALMKEPSIRH--LVVSGGVANSQYRARLDVVKKQGLQVLC 364
QY 176 PPRLCTDNGIMIAMNGIERLRAGIGILHDIEGIRYEPKCP 216
DB 365 PPRLCTDNGVMIAMTGIHFRMG-----RYDPPPP 395

RESULT 14
US-11-097-143-34191
Sequence 34191, Application US/11097143
Publication No. US20050208558A1
GENERAL INFORMATION:
APPLICANT: Venter, J. Craig
APPLICANT: et al.
TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
FILE REFERENCE: C1000728
CURRENT APPLICATION NUMBER: US/11/097,143
CURRENT FILING DATE: 2005-04-04
PRIOR APPLICATION NUMBER: 60/157,832
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: 60/160,191
PRIOR FILING DATE: 1999-10-19
PRIOR APPLICATION NUMBER: 60/161,932
PRIOR FILING DATE: 1999-10-28
PRIOR APPLICATION NUMBER: 60/164,769
PRIOR FILING DATE: 1999-11-12
PRIOR APPLICATION NUMBER: 60/173,383
PRIOR FILING DATE: 1999-12-28
PRIOR APPLICATION NUMBER: 60/175,693
PRIOR FILING DATE: 2000-01-12
PRIOR APPLICATION NUMBER: 60/184,831
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: 60/191,637
PRIOR FILING DATE: 2000-03-23
NUMBER OF SEQ ID NOS: 43008
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 34191
LENGTH: 409
TYPE: PRT
ORGANISM: DROSOPHILA
US-11-097-143-34191

Query Match      26.8%; Score 332.5; DB 20; Length 409;
Best Local Similarity 35.9%; Pred. No. 2.5e-27;
Matches 83; Conservative 38; Mismatches 97; Indels 13; Gaps 5;
```

```
QY 121 HLVKTRRAILFC--KQDRLPQNNNAVLSGGVANSFYIRALBITLTAQCTLLCPPP 178
DB 287 HLMHRTQRAIEYCLPBRQLFGDTPPTLVMSGGVANNDAIVANIEHLAAQYGRSFRPSK 346
QY 179 RCTDNGIMIAMNGIERLRAGIGILHDI-GRYIEKPCGLVDISVEGEA 228
DB 347 RYCSBNGVMIAMHGEVOL-----LQDKKASTRYDVL---SIDIQSAGPA 388

RESULT 15
US-10-067-443-3
Sequence 3, Application US/10067443
Publication No. US20030082782A1
GENERAL INFORMATION:
APPLICANT: Bristol-Myers Squibb Company
TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED I
FILE REFERENCE: D0073 NP
CURRENT APPLICATION NUMBER: US/10/067,443
CURRENT FILING DATE: 2002-02-05
PRIOR APPLICATION NUMBER: US 60/266,518
PRIOR FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: US 60/282,814
PRIOR FILING DATE: 2001-04-10
NUMBER OF SEQ ID NOS: 71
SOFTWARE: PatentIn version 3.0
SEQ ID NO 3
LENGTH: 463
TYPE: PRT
ORGANISM: Arabidopsis thaliana
US-10-067-443-3

Query Match      26.3%; Score 326; DB 14; Length 463;
Best Local Similarity 33.6%; Pred. No. 1.5e-26;
Matches 82; Conservative 34; Mismatches 84; Indels 44; Gaps 5;
```

```
QY 1 LVALVGVSDPFLILGKSLDIAPGMDLVKVARSLIKHPECSTMSGKAIETHAKQGNRF 60
DB 221 LVLAKHKLQYIQLGTTTDDAIGEAIDYTKAKMGLDMH-----RSGGPAVEELALBGDAK 275
QY 61 HFDIKPPLHAKNCDSEFTGLQHVTDKIIMKEKEGIEKQILSSAADIATVQHTMAC 120
DB 276 SVKFNVPKMKHKDCNFSYAGLKTQVRLAIEAKE-----IRNRADIAASFQVAVL 325
QY 121 HLVKTRRAILFCQKQDLPQNNNAVLSGGVANSFYIRALBITLTAQCTLLCPPRL 180
DB 326 HLEEKCERAIDWALE--LEPSIKHNVISGGVANSKYVRLNNIVENKNKLKVCPPSL 382
QY 181 CTDNGIMIAMNGIERLRAGIGILHDIEGIRY-----PKCPGLVDISKE 224
DB 383 CTDNGVMIAMTGIHFRVG-----RYDPPPAITEPDYVYDLRPMPLGEETVAKG 432
QY 225 VGEA 228
DB 433 RSEA 436
```

Search completed: November 10, 2005, 16:58:52
Job time : 64.1272 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: November 10, 2005, 16:52:13 ; Search time 4714.79 Seconds
(without alignments)
2456.267 Million cell updates/sec

Title: US-10-649-273-2_COPY_176_414
Perfect score: 1240
Sequence: 1 LLLLVGVSDFLILGKSLDI..... DISKVGASIKVPLAKMEI 239

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 2427607955 residues
Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+ p2n model -DEV=xlp
-Q=/cgr2/USPRO.spool/p/US10649273/runat_02112005_091338_15550/app.query.fasta.1.1429
-DB=GenBdb1 -QPM=faatap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blsum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pic -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USPR=US10649273 @CNC 1.1 7742 @runat_02112005_091338_15550 -NCPU=6 -ICPU=3
-NO_MMAP -LARGESUBTRY -NEG_SCORES=0 -WAIT -DSPBLCK=100 -LONGLOC
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:*
1: gb_ha:*
2: gb_hcg:*
3: gb_in:*
4: gb_cm:*
5: gb_ov:*
6: gb_pat:*
7: gb_pn:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_vl:*
14: gb_vl:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1240	100.0	1416	AR428809	Sequence
2	1240	100.0	1526	AR428809	Sequence
3	1240	100.0	1908	BC011904	Homo sapi
4	1240	100.0	2197	AR428803	Sequence

5	1213.5	97.9	1387	6	AR428808	Sequence
6	1213.5	97.9	1387	9	HS4295148	Sequence
7	1203	1245	1245	6	AX664697	Sequence
8	1203	97.0	1820	6	AX664695	Sequence
9	1086	87.6	1844	10	BC058172	Mus muscu
10	1079	87.0	1017	10	BC038910	Mus muscu
11	1059	85.4	2208	6	AX713716	Sequence
12	1059	85.4	2208	9	AK055441	Homo sapi
13	983	79.3	1546	10	BC078974	Rattus no
14	944.5	76.2	1109	10	BC051211	Mus muscu
15	809	65.2	860	5	BX930694	Gallus ga
16	806	65.0	1522	5	BX934991	Gallus ga
17	803	64.8	1558	6	BX930963	Gallus ga
18	725	58.5	14364	6	AR428807	Sequence
19	725	58.5	84115	9	AC013468	Homo sapi
20	666	53.7	249601	2	AC114153	Rattus no
21	666	53.7	308652	2	AC121478	Rattus no
22	646.5	52.1	256751	10	AC122925	Mus muscu
23	610	49.2	875	6	CQ721898	Sequence
24	384.5	31.0	1474	3	AK113378	Clona int
25	369.5	29.8	121251	5	AL591593	zebrafish
26	356	28.7	20021	10	AF367967	Mus muscu
27	356	28.7	179252	10	AF131205	Mus muscu
28	345	27.8	117322	5	AL672217	zebrafish
29	332.5	26.8	1576	3	AY051882	Drosophila
30	332.5	26.8	1601	6	CQ606432	Sequence
31	332.5	26.8	3656	6	CQ606431	Sequence
32	332.5	26.8	14679	2	AC018262	Drosophila
33	332.5	26.8	180263	3	AC010671	Drosophila
34	332.5	26.8	207432	3	AE003513	Drosophila
35	324	26.1	1443	8	AY024338	Arabidops
36	324	26.1	1474	8	AY117283	Arabidops
37	324	26.1	1567	8	AY063864	Arabidops
38	322	25.0	1557	8	AY084577	Arabidops
39	311.5	25.1	1672	8	AK070912	Oryza sat
40	305	24.6	110000	2	AP006501	Continuation (9 of
41	262.5	21.2	333800	1	SM591792	Sinorhizo
42	260	20.7	1718	8	AK099965	Oryza sat
43	257	20.7	349354	1	BX640416	Bordelell
44	254	20.5	260271	1	AE017258	Molbachia
45	253	20.4	349116	1	AP003003	Mesornitho

ALIGNMENTS

RESULT 1
AR428809
LOCUS AR428809 1416 bp DNA linear PAT 08-OCT-2004
DEFINITION Sequence 177 from patent US 6743619.
ACCESSION AR428809
VERSION AR428809.1 GI:53934009
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1416)
AUTHORS Tang, Y.T., Zhou, P., Goodrich, R., Liu, C., Asundi, V., Ren, F.,
Zhang, J., Zhao, Q.A., Yang, Y., Xue, A.J., Wehrman, T., Wang, J.-R.,
Wang, D., and Drmanac, R.T.
TITLE Nucleic acids and polypeptides
JOURNAL Patent: US 6743619-A 177 01-JUN-2004;
FEATURES
location/Qualifiers
source
1..1416
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN

Alignment Scores: 2,24e-113 Length: 1416
Pred. No.: 1240.00 Matches: 239
Score: 1240.00 Conservative: 0
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 100.00% Indels: 0
Query Match: 100.00%

DB: 6 Gaps: 0

US-10-649-273-2_COPY_176_414 (1-239) x AR428809 (1-1416)

QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20

DB 586 CTGTGGCATTGATGTTCAAGAGGTTTCAGATTTTCGCTTGTGAAAGCTTTTGACATTA 645

QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40

DB 646 GCACCAAGTGCACATGCTTGACAGAGTGCAGAAAGACTTTCTTTAATTAATCAATCCAGAG 705

QY 41 CysSerTherMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60

DB 706 TGTCTCCACCATGAGTGGTGGAAAGCCATAGAACATTTGGCCAAACAGAAATATGATTT 765

QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCyAspPheSerPheThrGly 80

DB 766 CATTTTGCATCAAAACCTCCCTTGATCATGCTAAATTTGTGATTTTCTTTTACTGGA 825

QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGluGlyIleGluLys 100

DB 826 CTTCAACAGCTTACTGATTAATTAATTAATGAAGAAAGAAAGAGAGATTTGAGAAAG 885

QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleAlaIleValGlnHisThrMetAlaCys 120

DB 886 GGGCAATTCCTGCTTGACAGACAGCATTTGCTGCACAGTACAGACACATGCGATGT 945

QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140

DB 946 CATCTTGAAAGAAACATCGGGCTATTCTGTTTGTAAAGCAGAGACTTGTACTCT 1005

QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaAspPheThrIleArgArg 160

DB 1006 CAAATATATGAGATCTGTGATCTGTGATCTGTGATCTGTGATCTGTGATCTGTGATCT 1065

QY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeu 180

DB 1066 GCTCTGGAATTTTAAACAGACAGACATGCTTGTGTGTCTCTCTCCAGACTA 1125

QY 181 CysThrAspAsnGlyIleMetIleAlaIleArgArgGlyIleGluArgLeuArgAlaGlyLeu 200

DB 1126 TGCACGTATATGAGCATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1185

QY 201 GlyIleLeuHisAspIleGluGlyIleArgGlyArgLysCysProLeuGlyValAlaAsp 220

DB 1186 GGCATTTTTAAACATGAGAGGATCCGCTATGAAACCAAAATGTCTCTTGAGATGAGC 1245

QY 221 IleSerLysGluValGlyLysAlaSerIleLysValProGlnLeuLysMetGluIle 239

DB 1246 ATATCAAAAGAGTTGGAGAGCTTCCATTAAGTACCAATTAATAAATGAGATTA 1302

RESULT 2

AR428809 1526 bp DNA linear PAT 18-DEC-2003

LOCUS Sequence 23 from patent US 6642041.

ACCESSION AR428809

VERSION AR428809.1 GI:40188595

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 1526)

AUTHORS Chen,J., Feder,J.N., Nelson,T.C., Krystek,S.R. and Duclos,F.

TITLE Polynucleotides encoding a novel metalloprotease, MP-1

JOURNAL Patent: US 6642041-A 23 04-NOV-2003;

FEATURES

source location/Qualifiers

1..1526

/organism="unknown"

/mol_type="genomic DNA"

ORIGIN

Alignment Scores:

Pred. No.: 2,456-113

Score: 1240.00

Percent Similarity: 100.00%

Best Local Similarity: 100.00%

Query Match: 100.00%

DB: 6 Gaps: 0

US-10-649-273-2_COPY_176_414 (1-239) x AR428809 (1-1526)

QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20

DB 85 CTGTGGCATTGATGTTCAAGAGGTTTCAGATTTTCGCTTGTGAAAGCTTTTGACATTA 144

QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40

DB 145 GCACCAAGTGCACATGCTTGACAGAGTGCAGAAAGACTTTCTTTAATTAATCAATCCAGAG 204

QY 41 CysSerTherMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60

DB 205 TGTCTCCACCATGAGTGGTGGAAAGCCATAGAACATTTGGCCAAACAGAAATATGATTT 264

QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCyAspPheSerPheThrGly 80

DB 265 CATTTTGCATCAAAACCTCCCTTGATCATGCTAAATTTGTGATTTTCTTTTACTGGA 324

QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGluGlyIleGluLys 100

DB 325 CTTCAACAGCTTACTGATTAATTAATTAATGAAGAAAGAAAGAGAGATTTGAGAAAG 384

QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleAlaIleValGlnHisThrMetAlaCys 120

DB 385 GGGCAATTCCTGCTTGACAGACAGCATTTGCTGCACAGTACAGACACATGCGATGT 444

QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140

DB 445 CATCTTGAAAGAAACATCGGGCTATTCTGTTTGTAAAGCAGAGACTTGTACTCT 504

QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaAspPheThrIleArgArg 160

DB 505 CAAATATATGAGATCTGTGATCTGTGATCTGTGATCTGTGATCTGTGATCTGTGATCT 564

QY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeu 180

DB 565 GCTCTGGAATTTTAAACAGACAGACATGCTTGTGTGTCTCTCTCCAGACTA 624

QY 181 CysThrAspAsnGlyIleMetIleAlaIleArgArgGlyIleGluArgLeuArgAlaGlyLeu 200

DB 625 TGCACGTATATGAGCATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 684

QY 201 GlyIleLeuHisAspIleGluGlyIleArgGlyArgLysCysProLeuGlyValAlaAsp 220

DB 685 GGCATTTTTAAACATGAGAGGATCCGCTATGAAACCAAAATGTCTCTTGAGATGAGC 744

QY 221 IleSerLysGluValGlyLysAlaSerIleLysValProGlnLeuLysMetGluIle 239

DB 745 ATATCAAAAGAGTTGGAGAGCTTCCATTAAGTACCAATTAATAAATGAGATTA 801

RESULT 3

BC011904 1908 bp mRNA linear PRI 23-DEC-2003

LOCUS Homo sapiens O-sialoglycoprotein endopeptidase-like 1, mRNA (cDNA

DEFINITION clone MGC:20293 IMAGE:4121450), complete cds.

ACCESSION BC011904

VERSION BC011904.2 GI:40225818

KEYWORDS MGC.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1908)

AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

Strausberg,R.L., Feingold,B.A., Grouse,L.H., Derge,J.G.,

Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,

Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,

Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marinina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stappleton, M., Soares, M.B., Bonald, M.P., Casavant, T.L., Schaefer, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loguigliano, N.A., Peters, G.J., Abramson, R.D., Mullahy, S.J., Bosak, S.A., McWhan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Viallay, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Pahey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butlerfield, Y.S., Krzywinski, M.I., Skalek, U., Smalins, D.E., Schermer, A., Schein, J.E., Jones, S.J., and Marra, M.A.

Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

12477932

2 (bases 1 to 1908)

Strausberg, R.

Direct Submission

Submitted (30-JUL-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NHG-MGC Project URL: <http://mgc.nci.nih.gov>

On Dec 19, 2003 this sequence version replaced gi:15080281.

Contact: MGC help desk

Email: gcgaps-remail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)

DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),

Galtherdsburg, Maryland;

Web site: <http://www.nisc.nih.gov/>

Contact: nisc.mgc@nih.gov

Ahter, N., Ayala, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakeley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, D., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, B., Kwong, P., Latic, P., Legaspi, R., Maduro, Q.L., Masello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Stantiripod, S., Thomas, P.J., Touchman, J.W., Tsurgon, C., Vogt, J.L., Walker, M.A., Weherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LIML at: <http://image.llnl.gov>

Series: IRAL Plate: 28 Row: 1 Column: 22.

Location/Qualifiers

1. 1908

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="MGC:20293 IMAGE:4121450"

/tissue_type="Muscle, rhabdomyosarcoma"

/clone_id="NIH_MGC_17"

/lab_host="DH10B-R"

/note="Vector: pOTB7"

1. 1908

/gene="OSGEP1"

/db_xref="LocusID:64172"

104. 1348

/gene="OSGEP1"

/codon_start=1

/product="OSGEP1 protein"

/protein_id="AAH1904.1"

/db_xref="GI:15080282"

/db_xref="LocusID:64172"

/translation="MLILITKAGVFPKPKRKYVEPLRSFNFHGTLLFLAKIVLGIET
SCDDTAAYVDETGNVGEAIVSGTEVHLTKGIVPPAQQIHEENIORIVQBALASAS
GVSPDSLSAIAITTIKPGIALSLGVLSPSLQVUQGLKKPFIPIHMEHAIITLITLNK

misc_feature

215. 1111

/gene="OSGEP1"

/note="Peptidase M22; Region: Glycoprotease family"

/db_xref="CD:PFam00814"

ORIGIN

Alignment Scores:

Pred. No.: 3,19e-113 Length: 1908

Score: 1240.00 Matches: 239

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 100.00% Indels: 0

DB: 9 Gaps: 0

US-10-649-273-2_COPY_176_414 (1-239) x BC011904 (1-1908)

QY 1 LeuLuuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAsp1le 20

DB 629 CTTGTCACATTAAGTTCACAGAGTTTCAGATTCTCTTGGAAAGCTTTGGACATA 688

QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuLysHisProGlu 40

DB 689 GCACACAGGAGCATGCTTGCACAGGTCGCAAGAGCTTTCTTATTAACATCCAGAG 748

QY 41 CysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyLysArgPhe 60

DB 749 TCCTCCACCATGAGTGGTGGAAAGCATATGAAATTTGGCCAAACAGAAATAGATT 808

QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCyAspPheSerPheThrGly 80

DB 809 CATTTTGACATCAAACTCCCTTGCATCATGCTTAAATAATGATTTCTTTTACTGGA 868

QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGlyIleGluLys 100

DB 869 CTTCAACAGCTTACATGATTAATTAATGAAAAGAAAAGAGAGCTATTGGAAG 928

QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 120

DB 929 GGGCAATCTGCTCTTACAGCAGACATGCTGCACAGTACAGACATGCGCATGT 988

QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140

DB 989 CATCTTGGAAAAAGAACATCGGGCTATTCTGTTTGTAGCAGAGACATGTTACCT 1048

QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheTyrlleArgArg 160

DB 1049 CAATAATTAAGCAGTCTGCTGTCATCTGGTGGTGGTGGCAAGTAACTTATATCCGACA 1108

QY 161 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180

DB 1109 GCTTGGAATTTTAAACAACCAACACAGTGCATCTTGTGTCTCTCCACAGACTA 1168

QY 181 CysThrAspAsnGlyIleMetIleAlaArgAsnGlyIleGluArgLeuArgAlaGlyLeu 200

DB 1169 TGCACATGATTAAGCATTAATGATTCATGGAATGGAATGAAGACTAGTCTGCTGGCTTG 1228

QY 201 GlyIleLeuHisAspIleGluGlyIleLeuArgTyrlleArgProLysCysProLeuGlyValAla 220

DB 1229 GGCATTTTAACATGACATAGAGAGCATCCGCTTGAACCAAAATGTCCTTGTAGATAGAC 1288

QY 221 HisSerLysGluValGlyGluAlaSerIleLysValAlaProGlnLeuLysMetGluIle 239

DB 1289 ATATCAAAAGAGATTGAGAGACCTTCATTAAGTACCAATTAAATTAATGAGATTA 1345

RESULT 4

AR428803 2197 bp DNA linear PAR 18-DEC-2003

LOCUS

DEFINITION Sequence 1 from patent US 6642041.

ACCESSION AR428803
VERSION AR428803.1 GI:40188589
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 2197)
AUTHORS Chen, J., Feder, J.N., Nelson, T.C., Krystek, S.R. and Duclos, F.
TITLE Polynucleotides encoding a novel metalloprotease, MP-1
JOURNAL Patent: US 6642041-A 1 04-NOV-2003;
FEATURES Location/Qualifiers
source 1..2197
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN

Alignment Scores:
Pred. No.: 3,77e-113 Length: 2197
Score: 1240.00 Matches: 239
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-649-273-2_COPY_176_414 (1-239) x AR428803 (1-2197)

QY 1 LeuLeuAlaLeuValGInGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20
DB 756 CTGTGGCATTTAGTTCAGAGAGTTTCAGATTTTCGCTTCTGGAAAGCTTTGGACATA 815

QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40
DB 816 GCACCAAGTGCAGCTTGCAGCAAGGTGCAGAAAGACTTTCTTATATAACATCCAGAG 875

QY 41 CysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60
DB 876 TGCCTCCACCATGAGTGTGGGAAAGCCATAGAACATTTGGCCAAACAGAAATGATTT 935

QY 61 HisPheAspIleLysProProLeuHisAlaLysAsnCysAspPheSerPheThrGly 80
DB 936 CATTTGACATCAACCTCCCTTGATCATGCTAAATAATGATTTTCTTTTACTGGA 995

QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGlyIleGluLys 100
DB 996 CTTCAACAGCTTATGATTAATAATATGAAAAAGAAAAAGAGAGTATTGAGAG 1055

QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleValIleGlnHisThrMetAlaCys 120
DB 1056 GGGCAAAATCCTGCTTCAGCAGACATGCTGCCACAGTACAGACACAATGCGCATGT 1115

QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140
DB 1116 CATCTTGTGAAMAACAACATCGGGCTATTCGTTTGTGAAGACAGAGACTGTGTAAGT 1175

QY 141 GlnAsnAsnAlaValIleValAlaSerGlyGlyValAlaSerAsnPheTyrIleArgArg 160
DB 1176 CAATAATATGACAGTACGCTGTCATCTGTGTGTCGCAAGTAACTTCTATATCCGACA 1235

QY 161 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeu 180
DB 1236 GCTCTGGAATTTTAAACAAACGCAACACAGTGCACCTTGTGTGTCTCTCCACAGACTA 1295

QY 181 CysThrAspAsnGlyIleMetIleAlaThrAsnGlyIleGluArgLeuArgAlaGlyLeu 200
DB 1296 TGCACATAATATGACATTAATATGTCATGCAAGTGAATGATTAAGAAAGCTAGCTGCTG 1355

QY 201 GlnIleLeuHisAspIleGluGlyIleArgTyrGlnProLysCysProLeuGlyValAsp 220
DB 1356 GGCATTTTACATGACATAGAAAGGCAATCCGCTATGAAACCAAAATGCTCTTGGAATGAC 1415

QY 221 IleSerLysGluValGlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 239
DB 1416 ATATCAAAAGAGATTGGAGAAAGCTTCATATAAAGTACCAATTAATAATGAGATA 1472

RESULT 5
AR428808 1387 bp DNA linear PAT 18-DEC-2003
LOCUS
DEFINITION Sequence 21 from patent US 6642041.
ACCESSION AR428808
VERSION AR428808.1 GI:40188594
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1387)
AUTHORS Chen, J., Feder, J.N., Nelson, T.C., Krystek, S.R. and Duclos, F.
TITLE Polynucleotides encoding a novel metalloprotease, MP-1
JOURNAL Patent: US 6642041-A 21 04-NOV-2003;
FEATURES Location/Qualifiers
source 1..1387
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN

Alignment Scores:
Pred. No.: 9,37e-111 Length: 1387
Score: 1213.50 Matches: 238
Percent Similarity: 90.15% Conservative: 0
Best Local Similarity: 90.15% Mismatches: 1
Query Match: 97.86% Indels: 25
DB: 6 Gaps: 1

US-10-649-273-2_COPY_176_414 (1-239) x AR428808 (1-1387)

QY 1 LeuLeuAlaLeuValGInGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20
DB 549 CTGTGGCATTTAGTTCAGAGAGTTTCAGATTTTCGCTTCTGGAAAGCTTTGGACATA 608

QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40
DB 609 GCACCAAGTGCAGCTTGCAGCAAGGTGCAGAAAGACTTTCTTATATAAACAATCCAGAG 668

QY 41 CysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60
DB 669 TGCCTCCACCATGAGTGTGGGAAAGCCATAGAGCAATTTGGCCAAACAGAAATGAGATT 728

QY 61 HisPheAspIleLysProProLeuHisAlaLysAsnCysAspPheSerPheThrGly 80
DB 729 CATTTGACATCAACCTCCCTTGATCATGCTAAATAATGATTTTCTTTTACTGGA 788

QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGlyIle----- 98
DB 789 CTTCAACAGCTTACTGATTAATAATATGAAAAAGAAAAAGAGAGTATATTCTTA 848

QY 98 ----- 98
DB 849 ATTAGTAAGTTGAACAGATTAATAATCTCGATGTCGCTTAATAATAGCTGCTCATTTTC 908

QY 99 -----GluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGln 115
DB 909 TGCAGTATAGAAAGGGGCAAAATCTGTCTTTCAGCAGCAGACATTTCTGCAACAGTACAG 968

QY 116 HisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGln 135
DB 969 CACACATGAGCATGTCATCTTGGAAGAAACACATCGGGCTATTCGTTTGTGAAGCAG 1028

QY 136 ArgAspLeuLeuProGlnAsnAsnAlaValIleValAlaSerGlyGlyValAlaSerAsn 155
DB 1029 AGAGACTGTACTCTCAAAATATATGACATGCTGTGTCATCTGTGTGTCGCAAGTAAC 1088

QY 156 PheTyrIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCys 175
DB 1089 TTCTATATCGCAGAGCTCTGGAATTTTAAACAAACGCAACACAGTGCACCTTGTGTGT 1148

QY 176 ProProProArgLeuCysThrAspAsnGlyIleMetIleAlaThrAsnGlyIleGluArg 195

Db 1149 CCTCCTCCAGACTATGATGATTAATGATTCATGATGATGATTAAGAA 1208
Qy 196 LeuargAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrgIleProLysCys 215
Db 1209 CTACGTGGTGGCTTGGGCAATTTTACATGACATAGAAAGCATCCGCTATGAACCAAAATGT 1268
Qy 216 ProLeuGlyValAspIleSerIleGlyValGlyIleGluIleValIleValIleProGlnIleu 235
Db 1269 CCTCTTGAGTAGACATATCAAAAGAGTGGAGAGCTTCATTAAGTACCAAAATTA 1328
Qy 236 LysMetGluIle 239
Db 1329 AAAATGGAGATA 1340
RESULT 6
HSA295148 1387 bp mRNA linear PRI 30-OCT-2000
LOCUS Homo sapiens mRNA for putative sialoglycoprotease type 2.
DEFINITION AJ295148
VERSION AJ295148.1 GI:11071726
KEYWORDS metallopeptidase; sialoglycoprotease.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1
AUTHORS Chen J.M., Fortunato M. and Barrett A.J.
TITLE Cloning and sequencing of a second human putative
sialoglycoprotease homologue
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 1387)
TITLE Chen J.M.
JOURNAL Direct Submission
Submitted (27-OCT-2000) Chen J.M., MRC Molecular Enzymology
Laboratory, The Babraham Institute, Babraham, Cambridge, CB2 4AT,
UNITED KINGDOM
FEATURES
source
1. .1387
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="uterus"
24. .1343
/function="metallopeptidase of family M22"
/codon_start=1
/product="putative sialoglycoprotease type 2"
/protein_id="CAC14666.1"
/db_xref="GI:11071727"
/db_xref="GOA:Q9H4B0"
/translation="MLITKTAGVFPKDSKGVYBPLSPFPHDETLPIHKIVGIET
SCDDTAAAVVDENGVGEAHSOTBVKTKGIVPPAQQGLHENTRIYQEAISAS
GVSPSDLSAIAITTKPGIALSLVGLSPSLQVQLKKPPIPIHMEHAALTRITNK
VEPFVLISGHCILALVGVSDPLILGSLDIAPDMDKVARRSLIKHEPCST
MSGGAIEHLAKGNRPHFDIKPDLHAKNDFSTGQIHYTDKIIMKEKEGFIIL
SKGQIINIPGLCKIAAHPCEYKGGIILSSAADIAATQVHMACGLVETRAILPFCK
ORDILPQNNVTVASGVASNFYRBALEITLITNAQCILLCPRLCTDNGIMAMNG
IERLRGGILIHDIIEGRIRPEKCPGLVDISKVBASIKVQLKMEI"
ORIGIN
Alignment Scores:
Pred. No.: 9.37e-111 Length: 1387
Score: 1213.50 Matches: 238
Percent Similarity: 90.15% Conservative: 0
Best Local Similarity: 90.15% Mismatches: 1
Query Match: 97.86% Indels: 25
Gaps: 1
US-10-649-273-2_COPY_176_414 (1-239) x HSA295148 (1-1387)
Qy 1 LeuLeuValLeuValGlnGlyValSerAspPheLeuLeuGlyIleSerLeuAspIle 20
Db 549 CTGTGGCATTAAGTTCAAGAGGTTTCAGATTTTCGCTTTCGAAAGCTTTGGACATA 608

Qy 21 AlaProGlyAspMetLeuAspIleValAlaArgIleLeuIleIleHisIleProGlu 40
Db 609 GCACACAGGTGACATGCTTACCAAGGTGGCAAAAGACTTTCTTAATAAAACATCCAGAG 668
Qy 41 CysSerThrMetSerGlyGlyIleValAlaIleGluHisLeuValIleGlnIleValAsnArgPhe 60
Db 669 TGTCTCACCATAGAGGTGGGAGAACCATAGACATTTGGCCAAACAGAAATAGATTT 728
Qy 61 HisPheAspIleLeuProProLeuHisHisAlaIleAsnCysAspPheSerPheThrGly 80
Db 729 CATTTTGACATCAAACTCCCTTGATCATGCTAAAAATTTGATTTTCTTTTACTGGA 788
Qy 81 LeuGlnHisValIleThrAspIleIleIleMetIleValGlyIleGluGlyIle----- 98
Db 789 CTTCAACACGTTACTGATTAATAATGAAAAAGAAAAAGAAAGAAAGATATATTCTA 848
Qy 98 ----- 98
Db 849 ATTAGTAAAGTTGAAACAGATTAATAATTCCTGGATTGTGCTAAATACTGCTCATTTTC 908
Qy 99 -----GluIleGlyGlnIleLeuSerSerAlaIleAspIleAlaIleThrValGln 115
Db 909 TGCAGGTATGAGAAAGGGGCAAAATCCTGTTTACAGACAGACATTTGCTCCACAGTACAG 968
Qy 116 HisThrMetAlaCysHisLeuValIleValArgThrHisArgAlaIleLeuPheCysIleGln 135
Db 969 CACACAAATGGCATGTCATCTTGTGAAGAAACACATCGGCTATTCTGTTTGTAAAGCAG 1028
Qy 136 ArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyValIleAsnArg 155
Db 1029 AGAGACTGTATCCCAAAATATATGACATGCTGTCATCTGTGTGTCCCAAGTAAAC 1088
Qy 156 PheTyrlleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThrIleuLeuCys 175
Db 1089 TTCATATCCGACAGCTCTGAAATTTTAAACAAACGCAACACAGTGCCTTGTGTGT 1148
Qy 176 ProProArgLeuCysThrAspAsnGlyIleMetIleAlaIleThrAsnGlyIleGluArg 195
Db 1149 CCTCCTCCAGCATATGACATGATATGATGATGATGATGATGATGATGATGATGATGATG 1208
Qy 196 LeuargAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrgIleProLysCys 215
Db 1209 CTACGTGGTGGCTTGGGCAATTTTACATGACATAGAAAGCATCCGCTATGAACCAAAATGT 1268
Qy 216 ProLeuGlyValAspIleSerIleGlyValGlyIleGluIleValIleValIleProGlnIleu 235
Db 1269 CCTCTTGAGTAGACATATCAAAAGAGTGGAGAGCTTCATTAAGTACCAAAATTA 1328
Qy 236 LysMetGluIle 239
Db 1329 AAAATGGAGATA 1340
RESULT 7
LOCUS AX664697 1245 bp DNA linear PAT 22-MAR-2003
DEFINITION Sequence 6 from Patent WO02074960.
VERSION AX664697
KEYWORDS AX664697.1 GI:29164457
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1
AUTHORS Leiby K.R., Kapeller-Libermann R. and Glucksmann M.
TITLE of human proteins and uses thereof
JOURNAL Patent: WO 02074960-A 6 26-SEP-2002;
Millennium Pharmaceuticals, Inc. (US)
FEATURES
source
1. .1245
/organism="Homo sapiens"

ORIGIN /mol_type="unassigned DNA"
/db_xref="taxon:9606"

Alignment Scores:

Pred. No.: 9.1e-110 Length: 1245
Score: 1203.00 Matches: 232
Percent Similarity: 98.33% Conservative: 3
Best Local Similarity: 97.07% Mismatches: 4
Query Match: 97.02% Indels: 0
Gaps: 0

US-10-649-273-2_COPY_176_414 (1-239) x AX664697 (1-1245)

QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20
Db CTGTGGCATAGTTTCAAGAGAGTTTCAGATTTTCCTTGGAAAGCTTTGGACATA 585
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40
Db GCACCAAGGTGACATGCTTGACAGAGGTGGCAAGAAAGACTTTCTTATATAAACATCCAGAG 645
QY 41 CysSerThrMetSerGlyGlyValAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60
Db TGCTCCACCATGAGTGTGGGAAAGCCATAGAACATTTGGCCAAACAAAGAAATAGATT 705
QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80
Db CATTTTGACATCAAACTCCCTTGACATGCTTAAATAATGTGATTTTCTTTTACTGGA 765
QY 81 LeuGlnHisValThrAspLysIleIleMetLysGlyGluGlyGluGlyIleGluLys 100
Db CTTCACACGTTTACGATTAATAATAGAAAACAGAAAACAAAGAAAGTATTGAGAAG 825
QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleAlaThrValGlnHisThrMetAlaCys 120
Db GGGCAAAATCTGTCTTCAGACAGACATGCTGCACAGATACACACAAATGGCATGT 885
QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuPro 140
Db CATCTTGAAAGAAACACATCGGGCTATCTGTTTGTAAAGACAGAACTGTACTCT 945
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAspPheThrIleArgArg 160
Db CAATAATATGACATGCTGTTGCACTGCTGTGTGCGAAGTAATTCTATATCCGACGA 1005
QY 161 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeu 180
Db GCTCTGGAAATTTTAAACAAACGCAACAGTGCATTTGTTGTCTCTCCACAGACTA 1065
QY 181 CysThrAspAsnGlyIleMetIleAlaThrPheAsnGlyIleGluArgLeuArgAlaGlyLeu 200
Db TGCACTGAAATGGCATTAATGATTCATGAAATGATTAAGAAAGACTGCTGCTGCTG 1125
QY 201 GlyIleLeuHisAspIleGluGlyIleArgIleArgIleProLysCysProLeuGlyValAsp 220
Db GGCAATTTTACATGACATAGAGAGCATCCGCTATGAAACAAATGTCTCTTGGATGAGC 1185
QY 221 HisSerLysGluValGlyGlnAlaSerIleLysValProGlnLeuLysMetGluIle 239
Db ATATCAAAAGAAAGTTGGAGAAGCTTCATTAAGTAACCAATTAATAATGAGATA 1242

RESULT 8
AX664695 1820 bp DNA linear PAT 22-MAR-2003
LOCUS AX664695
DEFINITION Sequence 4 from Patent WO02074960.
ACCESSION AX664695
VERSION AX664695.1 GI:29164455
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1
Leiby, K.R., Kapeller-Liebermann, R. and Glucksmann, M.
AUTHORS 38650, 28472, 5495, 65507, 81588 and 14354 methods and compositions
TITLE of human proteins and uses thereof
JOURNAL Patent: WO 02074960-A 4 26-SEP-2002;
Millennium Pharmaceuticals, Inc. (US)
location/Qualifiers

FEATURES

source 1..1820
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

CDS

146..1390
/note="unnamed protein product"
/codon_start=1
/protein_id="CAD80044.1"
/translation="MLILTKTGVFPFKSKRKYVEPLRSFNFPGTLFLKIVLGIET
/translation="MLILTKTGVFPFKSKRKYVEPLRSFNFPGTLFLKIVLGIET
SCDDRAAVDETGNVLGEATHSQFVHLKNGIYIPPAOOLHRENIQIYVDELAS
GVSPEDLSAIVATTIKPGLALSLGVLSFSLVQVGLKRFITIHMEARLITRLTNK
VEFPLVLLISGHCILALVQGSDFLLGSLDIPAGMDKVARLSLIRPECT
MSGKALHIAKQNRFPFDIKPPLHAKNCDPSTGLOHTDKNNENKQEGIEKG
OILSSADIAATVOHTMACILYKRTTHRAILFKORDLLPONNAVVAAGVSNFYIR
RALEITSAVTOCTTLCPPRLCTDNGIMIANNGIIRLRAGSLIHDIBGIRPEKPL
GVDISKEVSEISIKVPOLKMEI"

ORIGIN

Alignment Scores:

Pred. No.: 1.43e-109 Length: 1820
Score: 1203.00 Matches: 232
Percent Similarity: 98.33% Conservative: 3
Best Local Similarity: 97.07% Mismatches: 4
Query Match: 97.02% Indels: 0
Gaps: 0

US-10-649-273-2_COPY_176_414 (1-239) x AX664695 (1-1820)

QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20
Db CTGTGGCATAGTTTCAAGAGAGTTTCAGATTTTCCTTGGAAAGCTTTGGACATA 730
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40
Db GCACCAAGGTGACATGCTTGACAGAGGTGGCAAGAAAGCTTTCTTATATAAACATCCAGAG 790
QY 41 CysSerThrMetSerGlyGlyValAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60
Db TGCTCCACCATGAGTGTGGGAAAGCCATAGAACATTTGGCCAAACAAAGAAATAGATT 850
QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80
Db CATTTTGACATCAAACTCCCTTGACATGCTTAAATAATGTGATTTTCTTTTACTGGA 910
QY 81 LeuGlnHisValThrAspLysIleIleMetLysGlyGluGlyGluGlyIleGluLys 100
Db CTTCACACGTTTACGATTAATAATAGAAAACAGAAAACAAAGAAAGTATTGAGAAG 970
QY 911 CTTCAACAGTTTACGATTAATAATAGAAAACAGAAAACAAAGAAAGTATTGAGAAG 970
QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleAlaThrValGlnHisThrMetAlaCys 120
Db GGGCAAAATCTGTCTTCAGACAGACATGCTGCACAGTACAGACACAAATGGCATGT 1030
QY 971 GGGCAAAATCTGTCTTCAGACAGACATGCTGCACAGTACAGACACAAATGGCATGT 1030
QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuPro 140
Db CATCTTGAAAGAAACACATCGGGCTATCTGTTTGTAAAGAGAGACTGTTTACTCT 1090
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAspPheThrIleArgArg 160
Db CAATAATATGACATGCTGTTGCACTGCTGTGTGCGAAGTAATTCTATATCCGACGA 1150
QY 161 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeu 180
Db GCTCTGGAAATTTTAAACAAACGCAACAGTGCATTTGTTGTGCTCTCTCCACAGACTA 1210
QY 181 CysThrAspAsnGlyIleMetIleAlaThrPheAsnGlyIleGluArgLeuArgAlaGlyLeu 200

Db	1211	TGCACTGATTAATGCATTATGATTCATGCAATGATGATTTGAAGAAGCTACGTGCGCTTG	1270
Oy	201	GLY1LEUHI:saSPI:GLUGLY1LeaRgYrGUPLoLyCySProLeuGIYalasp	220
Db	1271	GGCAATTTACATGATGATGAAGGCGCTCGCTATGCAACCAAAATGCTCCCTTGAGTAGAC	1330
Oy	221	ILeserlySGIuValGIuGIuAlaSerIleIysValProGIInleuYMeRGIuile	239
Db	1331	ATATCAAAAGAAGTTGGAAAGCTTCATYAAAGTCAACATYAAATGAGATGA	1387
RESULT 9			
LOCUS	BC058172	1844 bp	linear
DEFINITION	Mus musculus cDNA clone MGC:67870 IMAGE:5012054, complete cds.		
ACCESSION	BC058172		
VERSION	BC058172.1		
KEYWORDS	MGC.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
REFERENCE			
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 1844)		
	Strausberg, R. L., Feingold, E. A., Grouse, L. H., Dergs, J. G., Klausner, R. D., Collins, F. S., Wagner, L., Shenmen, C. M., Schuler, G. D., Altschul, S. F., Zeeberg, B., Buetow, K. H., Schaefer, C. F., Bhat, N. K., Hopkins, R. F., Jordan, H., Moore, T., Max, S. I., Wang, J., Hsieh, F., DiCicco, L., Marusina, K., Farmer, A. A., Rubin, G. M., Hong, L., Stiplon, M., Soares, M. B., Bonaldi, A. M., Casavant, T. L., Scheetz, T. E., Brownstein, M. J., Udell, T. B., Toshiyuki, S., Carinci, P., Prange, C., Raha, S. S., Loquellano, N. A., Peters, G. J., Abramson, R. D., Mullahy, S. J., Bosak, S. A., McEwan, P. J., Wexler, K. J., Malek, J. A., Gunaratne, P. H., Richards, S., Wolter, K. C., Hale, S., Garcia, A. M., Gay, L. J., Hulyk, S. W., Valladon, D. K., Muzny, D. M., Sodergren, E. J., Lu, X., Gibbs, R. A., Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A. C., Shevchenko, Y., Bouffard, G. G., Blakesley, R. W., Touchman, J. W., Green, E. D., Dickson, M. C., Rodriguez, A. C., Grimwood, J., Schmutz, J., Myers, R. M., Butlerfield, Y. S., Krzywinski, M. I., Skalska, U., Smalls, D. E., Scherch, A., Schein, J. E., Jones, S. J., and Marra, M. A.		
TITLE	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A.	99 (26), 16899-16903	(2002)
MEDLINE	23288257		
PUBMED	12477932		
REFERENCE	2 (bases 1 to 1844)		
AUTHORS	Strausberg, R.		
TITLE	Direct Submission		
JOURNAL	Submitted (15-SEP-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA		
REMARK	NIH-MGC Project URL: http://mgc.ncl.nih.gov		
COMMENT	Contact: MGC help desk Email: cgabs-remail.nih.gov Tissue Procurement: Gilbert Smith, Ph.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINT) DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland; Web site: http://www.nisc.nih.gov/ Contact: nisc.mgc@nih.gov Ahter, N., Ayale, K., Beckstrom-Sternberg, S. M., Benjamin, B., Blakesley, R. W., Bouffard, G. G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N. L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Latic, P., Legaspi, R., Maduro, Q. L., Masiello, C., Maeser, B., Mastriani, J. C., McCabe, J. J., Pearson, R., Stanciro, S., Thomas, P. J., Touchman, J. W., Tsurgou, C., Vogt, J. L., Walker, M. A., Weetberry, K. D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E. D.		

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILNLT at: <http://image.llnl.gov>
Series: IRAX Plate: 123 Row: d Column: 22
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA bc1: 21332463.
Location/Qualifiers
1..1844
/organism="Mus musculus"
/mol_type="mRNA"
/strain="CZECH IT"
/db_xref="taxon:10090"
/clone="MGC:67870 IMAG:5012054"
/tissue_type="Mammary tumor metastasized to lung. Tumor arose spontaneously from a senescent normal mammary (clonal) outgrowth infected with the virus MMTV."
/clone_id="NCI_CGAP_Lu29"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
222..1466
/codon_start=1
/product="Unknown (protein for MGC:67870)"
/protein_id="AAH58172.1"
/db_xref="GI:34849664"
/translation="MLMRRTAGAIIPKPKSKYVFPDRFSVHPTLSCHLVYGIETSCDPTDAAVDENGVNFGALHSOTVHLKTGGIVPVAOILRENTORIVETLSASRIPTSDSAIATITIKPGLASLSTGGLSFSLQIVNRFKPKPIPHHMAHLITRLTKVEFFVLVLSGGHCLALVQGVDFLLGLSLDAFGDMIDVVARLULIKPECTMSGGAEIHLAKODNRPHFTINPQNAKCDSEFTSLQIHTDLITHKKEBGIENGQISSAADIAAAYOHATACHTAKRTHHAIFFCKQKNLSLPMNVLVSGVASLTYIRKALDIVANOTCTLCPPRLCTDNGIMIAWNGIERLRAGIAGVLHVEDIRYEPKCPGLGIDISREVAEAIKVPRLKNAL"
333..1397
/note="Q017; Region: Metal-dependent proteases with possible chaperone activity [posttranslational modification, protein turnover, chaperones]"
/db_xref="CDD:COG0533"

misc_feature
333..1397
/note="Q017; Region: Metal-dependent proteases with possible chaperone activity [posttranslational modification, protein turnover, chaperones]"
/db_xref="CDD:COG0533"

ORIGIN

Alignment Scores:
Pred. No.: 6.02e-98 Length: 1844
Score: 1086.00 Matches: 205
Percent Similarity: 92.05% Conservative: 15
Best Local Similarity: 85.77% Mismatches: 19
Query Match: 87.58% Indels: 0
DB: 10 Gaps: 0

US-10-649-273-2_COPY_176_414 (1-239) x BC058172 (1-1844)

Oy	1	LeuLeuAlaLeuValAlaGlnGlyValSerAspPheLeuLeuLeuGlyLeuSerLeuAspIle	20
Db	747	CTGTTGGCAATTAAGTCCAAAGGTGTTCCGATTTCTCGTCTCTTGAGAAAGTCTTGGACATA	806
Oy	21	AlAProGlyAspMetLeuAspLeuValAlaArgArgLeuSerLeuIleLeuShiSProGlu	40
Db	807	GGCGCAGGCGACATCGCTTGACAAAGGTGGCAAGAAAGCATTTCTTAATCAACATCCGAA	866
Oy	41	CysSerTherMetSerGlyGlyLeuValAlaIleGluHisLeuAlaLeuGlnGlyAsnArgPhe	60
Db	867	TGTTCTCAATAGATGGTGGAAAGCATATGAAACATTTGGCCAAAGACGGAATATGATTC	926
Oy	61	HisPheAspIleLeuProProGluLeuHisAlaAlaLeuAsnGlyAsnArgPheSerPheThGly	80
Db	927	CATTTCCTAATCAATCCACCTATGACAGAAATGCTAAGAAATGCGAATTTTCTTTTACGGGA	986
Oy	81	LeuGlnHisValThrAspLeuIleIleMetLeuLeuGlnGlyGluGluGluIleGluLeu	100
Db	987	CTTCAACATATTAATCTTAAGCTTAATACACACACAGAAAAGAAAGAAAGCATTTGAGAG	1046
Oy	101	GlyGlnIleLeuSerSerSerAlaAlaAspIleAlaAlaIleThyValGlnHisArgMetAlaCys	120
Db	1047	GGGCAATTCCTGTCATGACGGTGCAGACATTTGCTGTCGGGTAGACATGACAGACGCTGC	1106
Oy	121	HisLeuValLeuValArgThrHisAlaGluAlaIleLeuPheCysLeuGlnAlaArgLeuLeuPro	140

[illegible]

	BC Cancer Agency, Vancouver, BC, Canada
	info@bcsc.bc.ca
	Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
	Susanna Chan, Readman Chiu, Chris Fjell, Brin Garland, Ran Guin,
	Leticia Hsiao, Martin Krzywnski, Reta Kutsche, Oliver Lee, Soo
	Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
	Nees, Pawan Pandoh, Anna-Lisa Prabhu, Parvaneh Saeedi, Jacqueline
	Schein, Duane Smalius, Michael Smith, Lorraine Spence, Jeff Scott,
	Michael Thorne, Miranada Tsal, Nataja van den Bosch, Jill Vardy,
	George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E Consortium/LML at: <http://image.lml.gov>

Series: IRAX Plate: 86 Row: f Column: 12.
Location/Qualifiers
1. 1017
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5053559"
/issue_type="Liver, normal. 5 month old male mouse."
/clone_id="NCI CGAP L19"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"

ORIGIN

Alignment Scores:
Pred. No.: 1,48e-97 Length: 1017
Score: 1079.00 Matches: 205
Percent Similarity: 91.63% Mismatches: 14
Best Local Similarity: 85.77% Indels: 0
Query Match: 87.02% Gaps: 0

```
US-10-649-273-2_COPY_176_414 (1-239) x BC038910 (1-1017)
```

OY	1	leuleualaleuValgInglYvalSerAspPheleuleuleuglylYSerleuaSp1e	20
Db	110	CTGTGGGATTAAGTCACAAAGTGTTCCGATTTCCTCGTCCGGAAGTCCTTGACACTA	169
OY	21	AlaProglyaRpmMetleuaSryValAlaArgHrgleuSerleuIlElySH:SPROglu	40
Db	170	GCACCAGCGCAATGCTTGACAAGGTGGCAAGAAGACTTCTTTTAATCAACATCCAAA	229
OY	41	CysSerThrMetSerGlYglYlySalallegluH:isleualaYsgInglYasnArphe	60
Db	230	TGTTCTCAATGAAGTGGTGGAAAAAGCTATAGAACAGTTGGCCAAAGACGGAATTC	289
OY	61	HisPheaspIlElYsPROProleuHiHisAlaYsaAncYsaRPheSerPheThcglY	80
Db	290	CATTTCATCATTCACCATCACAAGATGCTAAAGATTGCCGATTTTTCTTTGACCGGA	349
OY	81	lenGlnH:ISvaLThrAspyIllelleMetLySgslySGlyugluglytleGluuu	100
Db	350	CTTPAACATATATCTGATNAGCTATAATACACACAAAGAAAAAGAGGCAATTGAGAG	409
OY	101	glYglnIleleuSerSerAlaAlaAspIlleaIalaThrvAlaGlnH:stHMeAlaCys	120
Db	410	GCGCAATTCCTGTCTATCAGCTGCAGACATTCCTCTCGCGTACAGCATGCAACGCGTGC	469
OY	121	HisleuVallySaRgTHrH:ISArgAlaalleudPheCylysGlnArGaspleuuePro	140
Db	470	CACCTTGGA AAAAACAACATCGCGCTATTCGTTTGGCAAGCAAGAAAAATTCCTCTCT	529
OY	141	GlnsnAsnAlaValleuValAlaSerGlyglYvalAlaSerAnpHeTyrlIlaArg	160
Db	530	CCAGCTTAACGAGATTAAGTTGATCTCGAAGSTYTGCAGATTAACCTGTACATCCGAAA	589
OY	161	AlaIeuGlnIleleuThrAsnAlaThcGlnCysThrleuIeuCysPROProArgIeu	180
Db	590	GCATTGGAAATTCGCAAAATGCAACGACAGTGCAGTTTGTGTCCACTCCAAAGACTG	649

QY 181 CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeu 200
Db 650 TGCACCTGACATGGCATCATGATTCAGATGAAATGAAATTAACCTGCTGGCTTG 709
QY 201 GlyIleLeuHisAspIleGluGlyIleArgTyrGluProLysCysProLeuGlyValAsp 220
Db 710 GGGCTTTACATGATGAGAAAGACATCGATATGACCAAAATGCTCTTGGAGTAGAC 769
QY 221 IleSerIleGluValGlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 239
Db 770 ATATCCAGAGAGAGTTCAGAAAGCTGCATTAAGTACCGCATTAATAATGCACTT 826
RESULT 11
AX713716 2208 bp DNA linear PAT 15-APR-2003
LOCUS Sequence 400 from Patent EP1293569.
DEFINITION AX713716
ACCESSION AX713716
VERSION AX713716.1 GI:29888642
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS
1 Isogai, T., Sugiyama, T., Otsuki, T., Wakamatsu, A., Sato, H., Ishii, S.,
Yamamoto, J. I., Isono, Y., Hio, Y., Otsuka, K., Nagai, K., Irie, R.,
Tamechika, I., Seki, N., Yoshikawa, T., Otsuka, M., Nagahari, K. and
Masuko, Y.
Full-length cDNAs
Patent: BP 1293569-A 400 19-MAR-2003;
Hella Research Institute (JF) ; Research Association for
Biotechnology (JF)
FEATURES
source 1. 2208
location/Qualifiers
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Alignment Scores:
Pred. No.: 3.57e-95 Length: 2208
Score: 1059.00 Matches: 211
Percent Similarity: 88.28% Conservative: 0
Best Local Similarity: 88.28% Mismatches: 4
Query Match: 85.40% Indels: 24
Gaps: 1
DB: 6
US-10-649-273-2_COPY_176_414 (1-239) x AX713716 (1-2208)
QY 1 LeuLeuAlaLeuValGlnGlyValaSerAspPheLeuLeuGlyLysSerLeuAspIle 20
Db 869 CTGTTGGCATTGTTCAAGAGATTTTCAGATTTCTGCTTGGAAAGTCTTTGACATA 928
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40
Db 929 GACACGAGTGAACGCTTGACAGAGTGCAGAGAGACTTCCTTAATTAACATCCAGAG 988
QY 41 CysSerThrMetSerGlyLysAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPhe 60
Db 989 TGCCTCCACCATGAGTGTGGGAAAGCATTTGACATTTGGCAACAGAAATAGATTT 1048
QY 61 HisPheAspIleLysProProLysHisAlaLysAsnCysAspPheSerPheThrGly 80
Db 1049 CATTTCACATCAAACTCCCTTCATCATGCTTAAATATGTAATTTCTTTTACCTGGA 1108
QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGluGlyIleGlyLys 100
Db 1109 CTTCAACACCTTACTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1168
QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnIleGlnMetAlaCys 120
Db 1169 GGGCAATATCTGCTTTCAGCAGAGACATTTGCTCCACAGTACGACACATATGATGT 1228

QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140
Db 1229 CATCTTGGAAAGAAACATCGGGCTATTCCTTTTGTAAACAGAGACATGTTACTT 1288
QY 141 GluAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheTyrIleArgArg 160
Db 1289 CAAATTAATGCGATGCTGCTTGCATCTGCTGCTGCGAAGTAACTTCTGTAATCCGACA 1348
QY 161 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180
Db 1349 GCTCGAAATTTTAACAAACGACACAGTSCACTTGTGTGCTCTCCACAGACTA 1408
QY 181 CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeu 200
Db 1409 TGCACCTGACATGGCATTAATGATTCGA-----TGATGCTCTTGGAGTAGAC 1435
QY 201 GlyIleLeuHisAspIleGluGlyIleArgTyrGluProLysCysProLeuGlyValAsp 220
Db 1436 -----TGATGCTCTTGGAGTAGAC 1456
QY 221 IleSerIleGluValGlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 239
Db 1457 ATATCAAAAGAGAGTTCAGAAAGCTTCATTAAGTACCAATTAATAATGAGACTA 1513
RESULT 12
AK055441 2208 bp mRNA linear PRI 30-JAN-2004
LOCUS Homo sapiens cDNA FLJ30879 f1s, clone FBEBA2004592, highly similar
DEFINITION to Homo sapiens mRNA for putative sialoglycoprotease type 2.
ACCESSION AK055441
VERSION AK055441.1 GI:16550166
KEYWORDS oligo capping; f1s (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS
1 Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R.,
Wakamatsu, A., Hayashi, K., Sato, H., Nagai, K., Kimura, K., Makita, H.,
Sekine, M., Ohayashi, M., Nishi, T., Shibahara, T., Tanaka, T.,
Ishii, S., Yamamoto, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y.,
Nagahari, K., Murakami, K., Yasuda, T., Iwayanagi, T., Wagatsuma, M.,
Shiratori, A., Sudo, H., Hosoiri, T., Kaku, Y., Kodaira, H., Kondo, H.,
Sugawara, M., Takahashi, M., Kanda, K., Yokoi, T., Furuya, T.,
Kikawa, E., Omura, Y., Abe, K., Kamihara, K., Katsuma, N., Sato, K.,
Tanikawa, M., Yamazaki, M., Ninomiya, K., Ishibashi, T., Yamashita, H.,
Murakawa, K., Fujimori, K., Tanai, H., Kimata, M., Watanabe, M.,
Hiraoka, S., Chiba, Y., Ishida, S., Ono, Y., Takiguchi, S., Watanabe, S.,
Yosida, M., Hotsuta, T., Kusano, J., Kanehori, K., Takahashi, Fujii, A.,
Hara, H., Tanase, T., Nomura, Y., Togiya, S., Komai, F., Hara, R.,
Takeuchi, K., Arita, M., Imose, N., Muesashino, K., Yuki, H., Oshima, A.,
Sasaki, N., Aotsuka, S., Yoshihawa, Y., Matsunawa, H., Ichihara, T.,
Shiohata, N., Sano, S., Moriya, S., Momiyama, H., Satoh, N., Takami, S.,
Terashima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H.,
Goto, Y., Shimizu, F., Wakebe, H., Hishigaki, H., Watanabe, T.,
Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K.,
Kumagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M.,
Tashiro, H., Tanigami, A., Fujisawa, T., Ono, T., Yamada, K., Fujii, Y.,
Ozaki, K., Hirao, M., Ohmori, Y., Kawabata, A., Hikiji, T., Kobatake, N.,
Inagaki, H., Ikema, Y., Okamoto, S., Okitani, R., Kawakami, T.,
Noguchi, S., Itoh, T., Shigeta, K., Senba, T., Matsumura, K.,
Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togaishi, T.,
Oyama, M., Hata, H., Watanabe, M., Komatsu, T., Mizushima-Shugano, J.,
Satoh, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K.,
Nagase, T., Nomura, N., Kikuchi, H., Masuko, Y., Yamashita, R.,
Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isogai, T. and Sugano, S.
Complete sequencing and characterization of 21,243 full-length
human cDNAs
Nat. Genet. 36 (1), 40-45 (2004)
14702039
JOURNAL
PUBMED
REFERENCE
AUTHORS
Nishi, T., Nakagawa, S., Senoh, A., Mizuguchi, H., Inagaki, H.,
Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S.,

Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Wagatsuma, M., Murakawa, K., Kanehori, K., Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahara, K., Masuko, Y., Nagai, K. and Isogai, T., NEDO human cDNA sequencing project

Unpublished
3 (bases 1 to 2208)
Isogai, T., Otsuki, T. and Sugiyama, T.
Direct Submission
Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: RAB and HRI.

FEATURES
source location/Qualifiers
1..2208
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="FEBRA2004592"
/tissue_type="brain"
/clone_id="FEBRA2"
/dev_stage="fetus"
/note="cloning vector: pME18SFL3"
344..1438
/note="unnamed protein product"
/codon_start=1
/protein_id="BAB70923.1"
/db_xref="GI:16550167"
/translation="MLITKTAGVFPFKRKRYEFLRSFNFHPTLFLKIVGLIET
GSDTAAVVDENGVGEAIIHSQTEVHLKTVLPAAOOLHENTORIYQALSLAS
VSPDLSAIVATTKTGALSTGYSRSLQVGLKKPKPIHMHMAHTITLTK
VERPFLVLTISGHCLALVQGVSDFLILGSLDIAFDMDKARPLIKHECST
MSGKAIENLAKQNPFPDIKPLHNAKNCDFPTGLOHTVDKIIMKEKEEESIEKS
OILSSADIATVOTMACHVKRTHRAIIFCKORDLIPONNAVIVASGVASNFICIR
RALEILTNAQCTILCPPRLCTDNGIMIA"

CDS
344..1438
/note="unnamed protein product"
/codon_start=1
/protein_id="BAB70923.1"
/db_xref="GI:16550167"
/translation="MLITKTAGVFPFKRKRYEFLRSFNFHPTLFLKIVGLIET
GSDTAAVVDENGVGEAIIHSQTEVHLKTVLPAAOOLHENTORIYQALSLAS
VSPDLSAIVATTKTGALSTGYSRSLQVGLKKPKPIHMHMAHTITLTK
VERPFLVLTISGHCLALVQGVSDFLILGSLDIAFDMDKARPLIKHECST
MSGKAIENLAKQNPFPDIKPLHNAKNCDFPTGLOHTVDKIIMKEKEEESIEKS
OILSSADIATVOTMACHVKRTHRAIIFCKORDLIPONNAVIVASGVASNFICIR
RALEILTNAQCTILCPPRLCTDNGIMIA"

ORIGIN
Alignment Scores:
Pred. No.: 3,57e-95 Length: 2208
Score: 1059.00 Matches: 211
Percent Similarity: 88.28% Conservative: 0
Best Local Similarity: 88.28% Mismatches: 4
Query Match: 85.40% Gaps: 24
DB: 9 Indels: 1
US-10-649-273-2_COPY_176_414 (1-239) x AK055441 (1-2208)

QY 1 LeuLeuAlaLeuValGInGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20
Db 869 CTGTTGGCATTAGTTCAAGAGGTTTCAGATTTCCTGTTTGAAGATCTTTGACACATA 928
QY 21 AlAProGlyAspMetLeuAspLysValAlaIArgArgLeuSerLeuIleLysHisProGlu 40
Db 929 GCACAGAGTACATCTTGACAGAGGTGACAGAGACCTTCCTTTATATAAATCATCAAG 988
QY 41 CysSerThrMetSerGlyLysAlaIleGluHisLeuAlaLysGInGlyAsnArgPhe 60
Db 989 TGCCTCCACCATGAGTGGGAGAAAGCCATAGAACATTTGGCCAAACAGAAATAGATT 1048
QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCyAspPheSerPheThrGly 80
Db 1049 CATTTCACATCAAAACCTCCCTTCATCATCTATAAAATTTGATTTCCTTTTACTGGA 1108
QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGluGlyIleGluLys 100

Db 1109 CTTCACACGCTTACTGATTAATAATATGAAAAAGAAAAAGAGAGATTGAGAG 1168
QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleValGlnHisThrMetAlaCys 120
Db 1169 GGGCAATCCGTCTTCAGCAGACAGACTTCCTGCCACAGTACAGACATGACATG 1228
QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140
Db 1229 CATCTTGAAAAAACAACATCGGGCTATCTGTTTGTAAAGCAGAGACTTTTACT 1288
QY 141 GlnAspAsnAlaValIleValAlaSerGlyLysValAlaSerAspPheThrIleArgArg 160
Db 1289 CAAATATATGACAGTACGTGTCATCTGTGTGTGCAAGTAACTTCATGATCCGACGA 1348
QY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180
Db 1349 GCTCTGAAATTTTACCAACGACACAGCGACTTCTGTGTCTCTCCACAGACTA 1408
QY 181 CysThrAspAsnGlyIleMetIleAlaThrAsnGlyIleGluArgLeuArgAlaGlyLeu 200
Db 1409 TGCACTGATATGACATATGATGCA----- 1435
QY 201 GlyIleLeuHisAspIleGluGlyIleArgGlyArgProLysCysProLeuGlyValAsp 220
Db 1436 -----TGATGCTCTTGGAGTAC 1456
QY 221 IleserLysGluValGlyLysIleLysValProGlnLeuLysMetGluIle 239
Db 1457 ATATCAAAAGAGTTGGAGAGCTTCATTAAGTACCAATTAAGAGAGATA 1513

RESULT 13
BC078974
LOCUS BC078974 1546 bp mRNA linear ROD 03-AUG-2004
DEFINITION Rattus norvegicus cDNA clone IMAGE:711906, partial cds.
ACCESSION BC078974
VERSION BC078974.1 GI:50926879
KEYWORDS
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 1546)
Straussberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Wax, S.I., Wang, J., Hsieh, F.,
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
Abramson, R.D., Mulhany, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richard, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hilyk, S.W.,
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriques, S.,
Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Boutfield, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalins, D.E.,
Scherer, A., Schein, J.R., Jones, S.J. and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
2 (bases 1 to 1546)
DIRECTOR MGC Project.
TITLE
AUTHORS
JOURNAL
REMARK
Submitted (02-AUG-2004) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>

COMMENT

Contact: MGC help desk
Email: cgabbs-remail.nih.gov
Tissue Procurement: Howard Jacobs
cDNA Library Preparation: Express Genomics
DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www-shgc.stanford.edu>
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>
Series: IRK Plate: 182 Row: f Column: 6
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, Similarity but not identity to protein.

FEATURES

Source

```
1..1546
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="IMAGE:7111906"
/tissue_type="Testis, rat (Brown Norway)"
/clone_lib="NIH MGC_237"
/lab_host="DH10B"
/notice="Vector: pExpress1"
144..1546
/codon_start=1
/product="unknown (protein for IMAGE:7111906)"
/protein_id="AAH78974.1"
/db_xref="GI:50926880"
/translation="MLMSKTAGATPRPRNVGRIRBNVOPALFHKKVIGET
SCDTRAAVDETGNVAGBALHSQTEVHLKTCGYPPAQQIHKRNIQRIVEALAS
GSPSDLSAIIATTKPGLALSDVGLSVQVNVNPKKFPPIHMEALITRLTHK
VGPPELVLSIGHCLALVQSDPLGLKSLDAPDMLDKVARRLSLIGPEST
MSGGVAIEHLAKENRFPHTINPMQNAKCPSTGLQHTVDKLIITHEKEGIEK
OILSSADIAAIVOHATACHLAKRTHRAIIFCOONLSPAAVLVSGEVSNNYIR
RALEIVNATGCTLCPPRLCTDNGIMIANNGIRLRLAGCGLIDVEDIRERPAGE
IVSRVATLTEDLDSVSRHTVAHSPINSGSRANIQTSMCCTQIVMTVRHRL
NINLKSRRKKKKKKKK"
```

CDS

Alignment Scores: 8.27e-88 Length: 1546
Score: 983.00 Matches: 188
Percent Similarity: 92.06% Conservative: 9
Best Local Similarity: 87.85% Mismatches: 17
Query Match: 79.27% Indels: 0
Gaps: 0

US-10-649-273-2_COPY_176_414 (1-239) x BC078974 (1-1546)

```
QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20
Db 669 CTGTTGGCGTTAGTTCAAGAGTCTTTCAGATTTCTGCTCTCGGAGAGCTCCGACATTA 728
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40
Db 729 GCGCAGGCGCATCTTGACAGAGTGCGAAGAACATTTCTTTATCATCAATCCAGAA 788
QY 41 CysSerThrMetSerGlyGlyValAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60
Db 789 TGTCTTCAATGAGTGGTGGGAAGACTATGAACTGATTTTCTTTTACGGGA 848
QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80
Db 849 CACTTACTATCAATCCACCCATGCAAGATGCTAAGAACTGATTTTCTTTTACGGGA 908
QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGluGlyIleGluLys 100
Db 909 CTTCAACATGTCAACCGATTAAGCTATTAACACACAGAAAAAGAAAGACATTAAGAG 968
```

ORIGIN

REMARK
COMMENT
Contact: MGC help desk
Email: cgabbs-remail.nih.gov
Tissue Procurement: Marcello Bento Soares, Ph.D.
cDNA Library Preparation: M. Bento Soares, University of Iowa
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbio.org>
contact: amadan@systemsbiology.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>
Series: IRK Plate: 113 Row: b Column: 1
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis.

FEATURES

Source

```
1..1109
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:1327545"
/tissue_type="Thymus gland, mouse"
/clone_lib="Soares_thymus_2NBMT"
/lab_host="DH10B"
/notice="Vector: p773D-Pac"
```

ORIGIN

```
QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleAlaThrValGlnHisThrMetAlaCys 120
Db 969 GGGCAATCCTGTCATCAGCCGACACATTCGTGCGGTACAGACGCAACAGCGTGC 1028
QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheLysGlnArgAspLeuLeuPro 140
Db 1029 CACTTCGGAAAAACACACATTCGTATTCCTTTTGGCAGCAAAAAATTTGCTATCT 1088
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheTyrlleArgArg 160
Db 1089 CCAGCTAACCGCATATTAGTTGTCTGTGAGAGGTTCGAAGTAACTTGATCATCCAGAA 1148
QY 161 AlaLeuGlnIleLeuThrAsnAlaArgGlnCysThrLeuLeuCysProProArgLeu 180
Db 1149 GCATTGGAATTTGACAAATGCAACACAAATGCAATGCAATTCCTGTTGTGTCCTCCGAGACTG 1208
QY 181 CysThrAspAsnGlyIleMetIleAlaIleArgGlyIleGluArgLeuAlaGlyLeu 200
Db 1209 TGCACCTGACAAATGATGATCATGATTCATGGAATGGAATTGAAGATTACCTGCTGCTTG 1268
QY 201 GlyIleLeuHisAspIleGluGlyIleArgTyrlleGluProLys 214
Db 1269 GGCATTTTACATGATGATGAGACATCCGATACGAACCAAG 1310
```

RESULT 14
BC051211 1109 bp mRNA linear ROD 15-Apr-2003
LOCUS
DEFINITION
Mus musculus, clone IMAGE:1327545, mRNA.
ACCESSION
BC051211 GI:29881634
VERSION
BC051211.1 GI:29881634
KEYWORDS
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1109)
Strausberg, R.
Direct Submission
Submitted (14-Apr-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk

REMARK

COMMENT


```

Qy 161 AlaleuGlulIleuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180
Db 574 GAGCTGCAGACTCTGGCAAAATGCAAACGGTTTGCTTTCTGTCTCTCTCCCAAGGCTG 633
Qy 181 CysThrAspAsnGlyIleMetIleAlaThrAsnGlyIleGluArgLeuArgAlaGlyLeu 200
Db 634 TGCACCGATATATGGGTATATGATTCATGAAATGGCATTGAAAGGTTGCGTCAGAGATGT 693
Qy 201 GlyIleLeuHisAspIleGluGlyIleArgTyrgIupProLysCysAsProLeuGlyValAsp 220
Db 694 GGATTTTATACAGTACTGATGGCATCCGCTACGAACCAAAAGCTCCCTTGGAATTGAT 753
Qy 221 ILeSerLysGluValGlyGluAlaSerIleLysValProGlnLeuLys 236
Db 754 ATTTCCAAAAGATTGAAGAGATTCCATCAAAAGTCCCAAGACTAAG 801

```

Search completed: November 10, 2005, 22:36:56
 Job time : 4721.79 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: November 10, 2005, 16:48:13 ; Search time 538.789 Seconds
(without alignments)
2625.922 Million cell updates/sec

Title: US-10-649-273-2_COPY_176_414
Perfect score: 1240
Sequence: 1 LIALVQGVSDFLILGKSLDI.....DISKVGESAIKVPQLKMEI 239

Scoring table: BLASTN62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4390206 segs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seg length: 0
Maximum DB seg length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+ p2n.model -DEV=x1d
-O=/cgn2_1/USFTO_spool_p/US10649273/runat_02112005_091337_15540/app_query.fasta_1.1429
-DB=N.Geneseq_16Dec04 -OPMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blonum62 -TRANS=human40.cdi
-LIST=45 -DOCLIN=200 -THR SCORE=PCT -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTPMT=p2n -NOM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10649273 @CGN 1.1 1063 @runat_02112005_091337_15540 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -MAIN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6
-Fgapext=7 -Ygapop=10 -Ygapext=0.5 -Delop=6 -Delext=7

Database : N.Geneseq_16Dec04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1240	100.0	1416	8	ABX70950
2	1240	100.0	1526	6	ABX70950
3	1240	100.0	2058	6	ABX70950
4	1240	100.0	2197	6	ABX70950
5	1240	100.0	2572	8	ABX70950

6	1203	97.0	1820	6	AA046856
7	1203	97.0	1820	8	ACA60887
8	1203	97.0	1821	10	ABX57020
9	1059	85.4	2208	10	ADA52832
10	1059	85.4	2890	12	ADQ24627
11	995.5	80.3	3358	10	ADBE3135
12	870	70.2	1572	6	ABQ75508
13	599	48.3	2734	5	AA846422
14	468	37.7	371	12	ADL86725
15	468	37.7	371	12	ADL86726
16	332.5	26.8	1601	4	AB124633
17	332.5	26.8	3656	4	AB124632
18	332	26.0	1557	3	ABC38454
19	257	20.7	1146	8	ACA26804
20	248.5	20.0	1000	6	AAE91424
21	248.5	20.0	1000	6	ABX37804
22	248.5	20.0	1044	8	ACA39102
23	248.5	20.0	94750	4	AAE28551
24	247	19.9	936	8	ACA20445
25	247	19.9	1053	12	ADL03120
26	241	19.4	1032	8	ACA43173
27	239	19.3	4360	6	AA048239
28	235.5	19.0	1092	6	ABQ90383
29	234	18.9	300	3	AAH00934
30	232	18.7	676	4	AAH08019
31	232	18.7	1385	4	AAH15110
32	231	18.6	1026	4	AA554064
33	231	18.6	1026	6	ACA42146
34	231	18.6	1026	10	ADG73341
35	231	18.6	1059	11	ABD02280
36	231	18.6	1206	11	ABD02197
37	229	18.5	1026	10	ADG73343
38	224	18.1	1029	4	AA553109
39	224	18.1	1029	8	ACA34150
40	224	18.1	11000	2	AA742063
41	217	17.5	9567	13	ADT05493
42	217	17.5	85814	13	ADT05644
43	214	17.3	1044	4	AB108591
44	213	17.2	1014	4	AA556045
45	213	17.2	1014	8	ACA51431

ALIGNMENTS

RESULT 1	ABX70950	standard; CDNA; 1416 BP.
ID	ABX70950	standard; CDNA; 1416 BP.
AC	ABX70950;	
DT	05-MAR-2003	(first entry)
XX	Novel human cDNA sequence #175.	
DE	Human; gene; ss; nervous system disorder; peripheral neuropathy;	
KW	Huntington's disease; amyotrophic lateral sclerosis; haemophilia;	
KW	neurodegenerative disease; Parkinson's disease; Alzheimer's disease;	
KW	autoimmune disease; systemic lupus erythematosus; rheumatoid arthritis;	
KW	insulin-dependent diabetes mellitus; anaemia; thrombocytopenia; wound;	
KW	ulcer; burn; bone disorder; osteoporosis; osteoarthritis; stroke;	
KW	fibrosis; reperfusion injury; infection; allergic rhinitis; asthma;	
KW	coagulation disorder; cancer; tumour; inflammatory disease; septic shock;	
KW	Crohn's disease; anaphylaxis; proliferation; chemotactic;	
KW	differentiation; stem cell growth factor; hematopoiesis; chemokinetic;	
KW	haemostatic; antiinflammatory; expressed sequence tag; EST.	
OS	Homo sapiens.	
PN	WO200281731-A2.	
XX	17-OCT-2002.	
PD	29-JAN-2002; 2002WO-US001222.	
XX		
PF		

XX 30-JAN-2001; 2001US-00774528.
XX (HYSE-) HYSEO INC.
XX (GOOD/) GOODRICH R. W.
XX Tang TY, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F,
XX Xue AJ, Yang Y, Wehrman T, Wang J, Wang D, Drmanac RT;
XX WPI; 2003-058563/05.
XX Novel polypeptide useful for treating neurodegenerative diseases, myeloid
XX or lymphoid cell disorders, bone disorders, mechanical and traumatic
XX disorders, coagulation disorders, and inflammatory diseases.
XX
XX Claim 1; Page; 612pp; English.
XX
XX This invention relates to the cDNA sequences encoding an isolated novel
XX human polypeptide. The protein encoded by the nucleic acid of the
XX invention is useful for treating central and peripheral nervous system
XX diseases (e.g. peripheral neuropathy, Huntington's disease, amyotrophic
XX lateral sclerosis); neurodegenerative diseases (e.g. Parkinson's disease,
XX Alzheimer's disease); autoimmune disease (e.g. systemic lupus
XX erythematosus, rheumatoid arthritis, insulin-dependent diabetes mellitus)
XX ; myeloid or lymphoid cell disorders (e.g. anaemia and thrombocytopenia)
XX ; wounds, ulcers, burns; bone disorders (e.g. osteoporosis,
XX osteoarthritis); mechanical and traumatic disorders (e.g. stroke, head
XX trauma); lung or liver fibrosis; reperfusion injury in various tissues;
XX bacterial, viral or fungal infections; allergic conditions such as
XX allergic rhinitis, asthma; coagulation disorders (e.g. haemophilia);
XX cancer and tumours; and inflammatory diseases (e.g. septic shock, Crohn's
XX disease, anaphylaxis). The protein may be used to inhibit the growth,
XX infection or function of infectious agents such as bacteria, fungi,
XX viruses, or to effect bodily characteristics, biorhythms or circadian
XX cycles of rhythms. The protein may also have
XX proliferation/differentiation, stem cell growth factor, haematopoiesis
XX regulation, immune stimulating or suppressing, chemotactic/chemokinetic,
XX haemostatic and thrombolytic, receptor/ligand, and antiinflammatory
XX activities. The cDNA sequences of the invention are useful for expressing
XX recombinant protein for analysis. The present sequence represents a novel
XX human cDNA sequence of the invention, this sequence is an expressed
XX sequence tag (EST) and was identified using subtractive hybridisation
XX
XX
SQ Sequence 1416 BP; 441 A; 280 C; 273 G; 422 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 3,21e-132 Length: 1416
Score: 1240.00 Matches: 239
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: Gaps: 0
US-10-649-273-2_copy_176_414 (1-239) x ABX70950 (1-1416)
QY 1 LeuLennalaLeuValGInglYValSerAspHeLeuLeuGluYLeSerLeuAspIle 20
DB 586 CTGTGGCATTAGTTCAGAGAGGATTTCGATTTTCGTTGGAAGTCTTTGGACATA 645
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40
DB 646 GCACCGAGTGAATGCTTGAACAAGTGGCAAGAACCTTTCTTTAATAAATCATCAAG 705
QY 41 CysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60
DB 706 TGCTCCACCATGAGTGGGGAAGCCATAGAACCTTTGGCCAAACAGAAATGATTT 765
QY 61 HisPheAspLysLeuProProLeuHisHisAlaLysAsnCyAspPheSerPheThrGly 80
DB 766 CATTTTACATCAAAACCTCCCTTCATCATCTAATAAATTTGATTTTCTTTTACTGGA 825
QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGluGlyIleGluLys 100

DB 826 CTTCAACAGCTTATCGATTAATAATATGAAAAAGAAAAAGACAGATTTGGAAG 885
QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleAlaThrValGlnHisThrMetAlaCys 120
DB 886 GGGCAAAATCTCTCTTCAGCAGACAGATGTGCGACAGTACAGACCAANTGGATGT 945
QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysValysGlnArgAspLeuPro 140
DB 946 CATCTTGGAAAAAGAACACATCGGCTATTCTGTTTGTAAACAGAGACATGTTACCT 1005
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAspPheThrIleArgArg 160
DB 1006 CAAAATATATGCGATCTGTTTCATCTGTGTGTGCGCAAGTAACCTTCTATATCCGACA 1065
QY 161 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeu 180
DB 1066 GCTCTGGAATTTTAAACCAACCCACAGTCATCTTTGTGTCTCTCCCAACATRA 1125
QY 181 CysThrAspAsnGlyIleMetIleAlaArgAsnGlyIleGluArgLeuArgAlaGlyLeu 200
DB 1126 TGCACTGATTAATGCAATTATGATTCATGGAATGGTAATGAAAGACTACGCTGCGCTTG 1185
QY 201 GlyIleLeuHisAspIleGluGlyIleArgGlyTrpGluProLysCysProLeuGlyValAsp 220
DB 1186 GGCATTTTACATGACATAGAAAGCATCCGCTATGAAACAAAATGTCCTCTTGAGATGAC 1245
QY 221 IleSerLysGluValGlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 239
DB 1246 ATATCAAAAGAAAGTTGGAGAGCTTCATTAAGTACCAATTAAATGGAGATA 1302
RESULT 2
ABS76639
ID ABS76639 standard; DNA; 1526 BP.
XX
XX ABS76639;
XX
XX 11-DEC-2002 (first entry)
XX
XX DNA encoding novel human metalloprotease MP1 fragment #1.
XX
XX
XX Metalloprotease, MP-1; immune disorder; glutamate transport; cancer;
XX motor neuron disorder; amyotrophic lateral sclerosis; ALS; diabetes;
XX reproductive disorder; Klinefelter's syndrome; germinal cell aplasia;
XX genital wart; metabolic disorder; premature puberty; Kallman syndrome;
XX Cushing's syndrome; neurodegenerative disease; Alzheimer's disease;
XX Parkinson's disease; Huntington's disease; Tourette syndrome; sepsis;
XX liver disease; renal disease; immune disorder; rheumatoid arthritis;
XX acquired immunodeficiency syndrome; AIDS; pulmonary disease; pneumonia;
XX emphysema; cystic fibrosis; vascular disorder; inflammatory disorder;
XX neurological disorder; gene; de.
XX
XX Homo sapiens.
XX
XX WO200272751-A2.
XX
XX 19-SEP-2002.
XX
XX 05-FEB-2002; 2002WO-US003353.
XX
XX 05-FEB-2001; 2001US-0266518P.
XX PR 10-APR-2001; 2001US-0282814P.
XX
XX (BRIM) BRISTOL-MYERS SQUIBB CO.
XX
XX Chen J, Feder J, Nelson TC, Duclos F, Krystek S;
XX P-PSDB; ABG96487.
XX WPI; 2002-723329/78.
XX
XX New isolated acid encoding MP-1 protein, useful for preventing,
XX treating, or ameliorating diseases associated with aberrant
XX metalloproteinase activity, e.g. immune, metabolic, inflammatory and
XX neurological disorders.
XX
XX

XX Disclosure; Page 462-463; 473pp; English.

XX The invention describes an isolated nucleic acid molecule (I) encoding a

CC metalloprotease (MP-1). (I) is useful for preventing, treating, or

CC ameliorating a medical condition, particularly an immune disorder, an

CC aberrant glutamate transport or motor neuron disorder, such as an

CC amyotrophic lateral sclerosis (ALS), its juvenile form or an ALS-like

CC condition. The compositions and methods are also useful for diagnosing,

CC prognosticating, treating, ameliorating and/or treating disorders

CC associated with MP-1 activity, e.g. diabetes, cancer, reproductive

CC disorders (e.g. Klinefelter's syndrome, genital warts, or germinal cell

CC aplasia), metabolic disorders (e.g. premature puberty, Kallman syndrome,

CC or Cushing's syndrome), neurodegenerative diseases (Alzheimer's disease,

CC Parkinson's disease, Huntington's disease or Tourette syndrome), liver

CC and renal diseases and immune disorders (e.g. AIDS, rheumatoid arthritis

CC or sepsis), pulmonary diseases (e.g. pneumonia, emphysema or cystic

CC fibrosis) and vascular, inflammatory and neurological disorders (e.g.

CC Alzheimer's disease or Parkinson's disease). This sequence represents a

CC metalloprotease MPI polynucleotide

XX

XX

Sequence 1526 BP; 484 A; 297 C; 289 G; 456 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	3,57e-132	Length:	1526
Score:	1240.00	Matches:	239
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-10-649-273-2_COPY_176_414 (1-239) x ABA93268 (1-1526)

QY 1 LeuenuAlaLeuValGlnGlyValSerApheLeuLeuGlyLysSerLeuasp11e 20

DB 85 CTGTTGCATTAAGTTCAGAGAGGTTTCAGATTTCCTTTCGAAAGCTTTGACATTA 144

QY 21 AlaproglyAspMetLeuAspLysValAlaArgArgLysSerLeuLysHisProGlu 40

DB 145 GCACCAAGTGCATGCTTGCAGACGTCGCAAGAAACCTTTCTTAATATAACATCCAG 204

QY 41 CysSerTherMetSerGlyGlyValAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60

DB 205 TGCTCCACATGATGCTGGGAAAGCCATAGAACATTTGGCCAAACAGAAATAGATT 264

QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCyAspPheSerPheThrGly 80

DB 265 CATTTCACATCAAAACCTCCCTTCATCATCTAATAAATGTGATTTTCTTTTACTGGA 324

QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGlnGluGlyIleGluLys 100

DB 325 CTTCACACGCTTACTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 384

QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 120

DB 385 GGGCAATTCCTGCTTCAGCAGACATTCCTGCAAGTACAGACACATGCGCATGT 444

QY 121 HisLeuValLysValGlnHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140

DB 445 CATCTTGAAAGAAACATCGGCTATTCGTGTTTGAAGCAGAGACTTTTACCT 504

QY 141 GluAsnAspAlaValLeuValAlaSerGlyGlyValAlaAspAspPheThrIleArgArg 160

DB 505 CAATAATATGACATGCTGTCATCTGCTGTCGCAAGTAACTTTATATCCGCGAGA 564

QY 161 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180

DB 565 GCTTGCAATTTTAACAAACGACACAGTCACTTTGTGTCTCTCCCGACATTA 624

QY 191 CysThrAspAsnGlyIleMetIleAlaIleArgGlnGlyIleGluArgLeuArgAlaGlyLeu 200

DB 625 TGCACTATATATGACATTAATGATTCAGTGAATGTATTAAGTAAAGCTAGTCTGCTTG 684

QY 201 GlyIleLeuHisAspIleGluGlyIleArgGlyGluProLysCysProLeuGlyValAsp 220

DB 685 GGCATTTTACATGACATAGAGGATCCGCTATGAACAAATGTCTCTTGAGTACAC 744

QY 221 HisSerLysGluValGlyGlnLysSerIleLysValProGluLeuLysMetGluIle 239

DB 745 ATATCAAAAGAAAGTTGAGAGAGCTTCATTAAGTACCAATTAATTAATTAATTA 801

RESULT 3

ABA93268

ID ABA93268 standard; cDNA; 2058 BP.

AC ABA93268;

DT 19-APR-2002 (first entry)

DE Human O-sialoglycoproteinase-like protein encoding cDNA SEQ ID NO:1.

XX Human; O-sialoglycoproteinase-like protein; OSGPLP; enzyme; gene; ss.

KW Homo sapiens.

OS Homo sapiens.

PH Key Location/Qualifiers

FT CDS /tag= a /product= "O-sialoglycoproteinase-like protein"

PN CN1318550-A.

PD 24-OCT-2001.

PP 19-APR-2000; 2000CN-00106834.

PR 19-APR-2000; 2000CN-00106834.

PA (SHAN-) SHANGHAI SHENGYUAN GENE DEV CO LTD.

PI Mao Y, Xie Y;

DR MPI: 2002-115090/16.

DR P-PSDB; ABB05481.

PT O-sialoglycoproteinase-like protein and encoding polynucleotide, useful

PT for diagnosing, preventing and treating related diseases.

XX Claim 5; Page 29-30 (Disclosure); 38pp; Chinese.

CC The present sequence encodes human O-sialoglycoproteinase-like protein

CC (OSGPLP). The present invention also describes: (1) the preparation of

CC the OSGPLP protein; (2) applying the OSGPLP protein in diagnosis; (3) the

CC prevention and/or treatment of related diseases; (4) utilizing the OSGPLP

CC protein in screening its agonist, excitomotor and inhibitor and preparing

CC an antibody against the OSGPLP protein; and (5) the use of the OSGPLP

CC polynucleotide sequences, proteins, agonists, excitomotor, inhibitors

CC and antibodies in treating diseases related to the abnormal OSGPLP gene

CC and in preparing the medicine composite for the treatment

XX

XX Sequence 2058 BP; 637 A; 400 C; 410 G; 611 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	5,45e-132	Length:	2058
Score:	1240.00	Matches:	239
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-10-649-273-2_COPY_176_414 (1-239) x ABA93268 (1-2058)

QY 1 LeuenuAlaLeuValGlnGlyValSerApheLeuLeuGlyLysSerLeuasp11e 20

DB 635 CTGTTGCATTAAGTTCAGAGAGTTCAGATTTCCTTTCGAAAGCTTTGACATTA 694

```
QY 21 AlaProGlyaspMetLeuaspLyseValAlaArgArgLeuSerLeuIleLyseHisProGlu 40
DB 695 GCACCAAGTGAACATGCTTGAACAAGTGGCAAGAAGACTTTCTTAATAATGAACATCCAG 754
QY 41 CysSerThrMetSerGlyGlyValAlaIleGluHisLeuAlaLyseGlnGlyAsnArgphe 60
DB 755 TGCTCCACCATGAATGGGGAAGCCCATGAAACCTTTGGCCAAACGAAGAAATGATTT 814
QY 61 HisPheAspIleLyseProProLeuHisHisAlaLyseAsnCyaspPheSerPheThrGly 80
DB 815 CATTTCACATCAAACTCCCTTCATCATCTAATAAATTTGATTTTCTTTTACTGA 874
QY 81 LeuGlnHisValThrAspLyseIleIleMetLyseGlyGlyGluGlyIleGluLyse 100
DB 875 CTTCACACCGTTACTGATTAATAATATGAAGAAAGAAAGAGATTTGAGAAG 934
QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 120
DB 935 GGGCAAAATCCGTCTTCAGACAGACATTCCTCCACAGTACAGACAAATGGCATGT 994
QY 121 HisLeuValLyseArgThrHisArgAlaIleLeuPheCysLyseGlnArgAspLeuPro 140
DB 995 CATCTTGTGAAAAGAACATCGGGCTATTCTGTTTGTAAAGCAGAGACTTTGACT 1054
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaAsnAsnPheTyrIleArgArg 160
DB 1055 CAAATATATGACGATACGTGGTTCATCTGTGGTGTGCAAGTAACTTCTATATCCGAGA 1114
QY 161 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeu 180
DB 1115 GCTCTGAAATTTTAAACAAACGCAACAGTGCATTGTGTGTGCTCCCTCCAGACTA 1174
QY 181 CysThrAspAsnGlyIleMetIleAlaThrAsnGlyIleGluArgLeuArgAlaGlyLeu 200
DB 1175 TGCACTATATATGATTAATATATGATGATGATGATGATGATGATGATGATGATG 1234
QY 201 GlyIleLeuHisAspIleGluGlyIleArgTyrGluProLyseCysProLeuGlyValAsp 220
DB 1235 GGCATTTTACATGACATGAAAGGATCCGCTATGAAACAAATGTCCTTGGATGAC 1294
QY 221 IleSerLyseGluValGlyGluAlaSerIleLyseValProGlnLeuLyseMetGluIle 239
DB 1295 ATATCAAAAGAAAGTGGAGAAGCTTCATTAAGTACCAATTAATAATGAGATA 1351
RESULT 4
ABST6635
ID ABST6635 standard; DNA; 2197 BP.
XX
XX ABST6635;
XX
DT 11-DEC-2002 (first entry)
DE DNA encoding novel human metalloprotease MPI.
XX
XX Metalloprotease; MP-1; immune disorder; glutamate transport; cancer;
XX motor neuron disorder; amyotrophic lateral sclerosis; ALS; diabetes;
XX reproductive disorder; Klinefelter's syndrome; germinal cell aplasia;
XX genital wart; metabolic disorder; premature puberty; Kallman syndrome;
XX Cushing's syndrome; neurodegenerative disease; Alzheimer's disease;
XX Parkinson's disease; Huntington's disease; Tourette syndrome; sepsis;
XX liver disease; renal disease; immune disorder; rheumatoid arthritis;
XX acquired immunodeficiency syndrome; AIDS; pulmonary disease; pneumonia;
XX emphysema; cystic fibrosis; vascular disorder; inflammatory disorder;
XX neurological disorder; gene; ds.
XX
XX Homo sapiens.
XX
XX WO200272751-A2.
XX
XX 19-SEP-2002.
XX
XX 05-FEB-2002; 2002WO-US003353.
XX
```

```
PR 05-FEB-2001; 2001US-0266518P.
PR 10-APR-2001; 2001US-0282814P.
XX
XX (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX
XX Chen J, Feder J, Nelson TC, Duclos F, Krystek S;
XX
XX MPI; 2002-723329/78.
XX
XX P-PSDB; ABG96478.
XX
XX New isolated nucleic acid encoding MP-1 protein, useful for preventing,
XX treating, or ameliorating diseases associated with aberrant
XX metalloproteinase activity, e.g. immune, metabolic, inflammatory and
XX neurological disorders.
XX
XX Claim 1, Fig 1A-C; 473pp; English.
XX
XX The invention describes an isolated nucleic acid molecule (I) encoding a
XX metalloprotease (MP-1). (I) is useful for preventing, treating, or
XX ameliorating a medical condition, particularly an immune disorder, an
XX aberrant glutamate transport or motor neuron disorder, such as
XX amyotrophic lateral sclerosis (ALS), its juvenile form or an ALS-like
XX condition. The compositions and methods are also useful for diagnosing,
XX prognosticating, treating, ameliorating and/or treating disorders
XX associated with MP-1 activity, e.g. diabetes, cancer, reproductive
XX disorders (e.g. Klinefelter's syndrome, genital warts, or germinal cell
XX aplasia), metabolic disorders (e.g. premature puberty, Kallman syndrome,
XX or Cushing's syndrome), neurodegenerative diseases (Alzheimer's disease,
XX Parkinson's disease, Huntington's disease or Tourette syndrome), liver
XX and renal diseases and immune disorders (e.g. AIDS, rheumatoid arthritis
XX or sepsis), pulmonary diseases (e.g. pneumonia, emphysema or cystic
XX fibrosis) and vascular, inflammatory and neurological disorders (e.g.
XX Alzheimer's disease or Parkinson's disease). This sequence represents a
XX metalloprotease MPI polynucleotide
XX
XX Sequence 2197 BP; 681 A; 441 C; 439 G; 636 T; 0 U; 0 Other;
XX
XX Alignment Scores:
XX Pred. No.: 5,98e-132 Length: 2197
XX Score: 1240.00 Matches: 239
XX Percent Similarity: 100.00% Conservative: 0
XX Best Local Similarity: 100.00% Mismatches: 0
XX Query Match: 100.00% Indels: 0
XX DB: Gaps: 0
XX
XX US-10-649-273-2_COPY_176_414 (1-239) x ABST6635 (1-2197)
QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLyseSerLeuAspIle 20
DB 756 CTGTGGCATTTAGTTCAAGAGATTTTCTGCTTCTTGAAAGCTTTGGACATA 815
QY 21 AlaProGlyAspMetLeuAspLyseValAlaArgArgLeuSerLeuIleLyseHisProGlu 40
DB 816 GCACCAAGTGAACATGCTTGAACAAGTGGCAAGAAGCTTTCTTAATAATGAACATCCAG 875
QY 41 CysSerThrMetSerGlyGlyValAlaIleGluHisLeuAlaLyseGlnGlyAsnArgphe 60
DB 876 TGCTCCACCATGAATGGGGAAGCCCATGAAACCTTTGGCCAAACGAAGAAATGATTT 935
QY 61 HisPheAspIleLyseProProLeuHisHisAlaLyseAsnCyaspPheSerPheThrGly 80
DB 936 CATTTCACATCAAACTCCCTTCATCATCTAATAAATTTGATTTTCTTTTACTGA 995
QY 81 LeuGlnHisValThrAspLyseIleIleMetLyseGlyGlyGluGlyIleGluLyse 100
DB 996 CTTCACACCGTTACTGATTAATAATATGAAGAAAGAAAGAGATTTGAGAAG 1055
QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 120
DB 1056 GGGCAAAATCCGTCTTCAGACAGACATTCCTCCACAGTACAGACAAATGGCATGT 1115
QY 121 HisLeuValLyseArgThrHisArgAlaIleLeuPheCysLyseGlnArgAspLeuPro 140
```

DB 1116 CATCTTGTAAGAAACACATCGGGCTATTCTGTTTGTAAAGACAGAGACTTGTTACTT 1175
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAspPheTyrIleArgArg 160
DB 1176 CAAATATATGACAGTACTGTTGATCTGGTGTCGCAAGTAACTTCTATATCCGAGCA 1235
QY 161 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrIleLeuLeuCysProProProArgLeu 180
DB 1236 GCGCTGGAAATTTTAAACAAACGCAACAGCGACTTGTGTGTCTCCCTCCCAAGACTA 1295
QY 181 CysThrAspAsnGlyIleMetIleAlaTyrAsnGlyIleGluValGluLeuArgAlaGlyLeu 200
DB 1296 TGCATGATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1355
QY 201 GlyIleLeuHisAspIleGluGlyIleArgTyrGluProLysCysProLeuGlyValAsp 220
DB 1356 GGCATTTTACATACATAGAGAGATCCGCTATGACCAAAATCTCTCTTGAGTACAC 1415
QY 221 ILeSerLysGluValGlyValAspSerIleLysValProGluLeuLysMetGluIle 239
DB 1416 ATATCAAAAGAAAGTGGAGAGCTTCATATAAGTACACAAATTAAATGAGATA 1472
RESULT 5
ABT23207
ID ABT23207 standard; DNA; 2572 BP.
XX ABT23207;
AC
XX
XX 01-MAY-2003 (first entry)
XX
XX Human protein modification + maintenance molecule DNA SEQ ID No 36.
XX
XX Cytostatic; antiarteriosclerotic; anti-HIV; anti-allergic; anticonvulsant;
XX
XX cerebroprotective; antiparkinsonian; nootropic; antiinflammatory;
XX
XX antitumor; hepatocytic; gynecological; antibacterial; virucide;
XX
XX protozoicide; antiparasitic; cell proliferative disease; PMOD;
XX
XX protein modification and maintenance molecule; immunogenic fragment;
XX
XX cancer; autoimmune; inflammatory disease; neurological disorder;
XX
XX gastrointestinal; developmental; vesicle trafficking disorder; infection;
XX
XX protein-protein interaction; drug-target interaction;
XX
XX gene expression profile; human; gene; ds.
XX
XX Homo sapiens.
XX
XX MO2003000844-A2.
XX
XX 03-JAN-2003.
XX
XX 18-JUN-2002; 2002MO-US019360.
XX
XX 22-JUN-2001; 2001US-0300508P.
XX
XX 06-JUL-2001; 2001US-0303445P.
XX
XX 13-JUL-2001; 2001US-0305405P.
XX
XX 09-AUG-2001; 2001US-031442P.
XX
XX 24-AUG-2001; 2001US-0314821P.
XX
XX 29-AUG-2001; 2001US-0315922P.
XX
XX 03-MAY-2002; 2002US-0378205P.
XX
XX (INCY-) INCYTE GENOMICS INC.
XX
XX
XX Gandhi AR, Kable AE, Swarnakar A, Hafalia AJA, Tran B, Duggan BM;
XX
XX Warren BA, Ison CH, Honchell CD, Nguyen DB, Lu DM, Lee EH, Yue H;
XX
XX Forsythe TJ, Barroso I, Ramkumar J, Griffin JA, Li JX, Yang J;
XX
XX Thangavelu K, Gietzen KJ, Ding L, Baughn MK, Borowsky ML, Yao MG;
XX
XX Walla NK, Mason PM, Gururajan R, Lee S, Becha SD, Lee SY, Tran UK;
XX
XX Elliott VS, Luo W, Sprague MW, Tang YT, Lu Y, Zebajadian Y;
XX
XX WPI: 184039/18.
XX
XX P-PSDB; ABU26654.
XX
XX New isolated human PMOD polypeptide and polynucleotide, useful for
XX
XX PT diagnosing, treating and preventing diseases or conditions associated
XX
XX PT with the aberrant PMOD expression, e.g. cancer, AIDS, atherosclerosis and

PT infections.
XX
XX Claim 91; Page 211; 225pp; English.
XX
XX The invention relates to an isolated polypeptide comprising: any of 28
XX
XX sequences of 48-1256 amino acids; a natural amino acid sequence at least
XX
XX 90% identical to the 28 amino acid sequences; 94% identical to a sequence
XX
XX of 703 or 267 amino acids, 96% identical to a sequence of 414 amino
XX
XX acids, or 97% identical to a sequence of 242 amino acids, all given in
XX
XX the specification; or a biologically active or immunogenic fragment of
XX
XX the isolated polypeptide. The polypeptides and polynucleotides are useful
XX
XX in diagnosing, treating and preventing diseases or conditions associated
XX
XX with the decreased expression of protein modification and maintenance
XX
XX molecules (PMOD), such as cell proliferative diseases (e.g. cancer,
XX
XX atherosclerosis), autoimmune/inflammatory diseases (e.g. AIDS,
XX
XX allergies), neurological disorders (e.g. stroke, Parkinson's disease,
XX
XX epilepsy), gastrointestinal (e.g. ulcer, cirrhosis), reproductive (e.g.
XX
XX endometriosis), developmental, vesicle trafficking disorders, and
XX
XX infections (e.g. bacterial, viral, parasitic, protozoal). These are also
XX
XX useful in assessing the effects of exogenous compounds on the expression
XX
XX of nucleic acid and amino acid sequences of PMOD. The PMOD or its
XX
XX fragments are useful in screening compounds for effectiveness as agonist
XX
XX or antagonist of the polypeptides, or in altering the expression of the
XX
XX target polynucleotide and compounds that specifically bind to or modulate
XX
XX the activity of the polypeptide. The microarray is useful in monitoring
XX
XX or measuring protein-protein interactions, drug-target interactions, and
XX
XX gene expression profiles. This polynucleotide sequence represents the DNA
XX
XX encoding a human PMOD protein of the invention
XX
XX
XX Sequence 2572 BP; 780 A; 489 C; 509 G; 794 T; 0 U; 0 Other;
XX
XX
XX Alignment Scores:
XX
XX Pred. No.: 7,47e-132 Length: 2572
XX
XX Score: 1240.00 Matches: 239
XX
XX Percent Similarity: 100.00% Conservative: 0
XX
XX Best Local Similarity: 100.00% Mismatches: 0
XX
XX Query Match: 100.00% Indels: 0
XX
XX DB: 8 Gaps: 0
XX
XX US-10-649-273-2_COPY_176_414 (1-239) x ABT23207 (1-2572)
QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20
DB 669 CTGTTGGCATTAAGTTCAAGAGAGTTTCAGATTTCCTGCTTGAAAGCTTTGGACATA 728
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgGluSerLeuIleLysHisProGlu 40
DB 729 GCACCAAGTGCATGCTTGCACAAAGTGCACAAAGACTTCTTAATAAATCATCCAGG 788
QY 41 CysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60
DB 789 TGTCTCACCATGAGTGGGGAAGCCATAGAACTTTGGCCAAACGAAATAGATT 848
QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80
DB 849 CATTTGACATCAAACTCCCTTGATCATCTAATAAATTGATTTTCTTTACTGCA 908
QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGlyIleGluLys 100
DB 909 CTTCAACAGCTTACTGATTAATAATATAAGAAAAAGAAAGAGAGATTGAGAAG 968
QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetLacys 120
DB 969 GGGCAAAATCCGTCTTCGACGACAACTTCCTGCACAGTACACACACAAATGCAAT 1028
QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGluArgAspLeuPro 140
DB 1029 CATCTTGAAAGAAACATCGGCTATTCTGTTTGTAAAGACAGACTTGTACTT 1088
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAspPheTyrIleArgArg 160
DB 1089 CAAATATATGACAGTACTGTTGATCTGGTGTCGCAAGTAACTTCTATATCCGAGCA 1148

QY 161 AlAlenGluIleLeuThraAsnAlaThrGlnCysThrIleuLeuCysProProProArgLeu 180
DB 1149 GCTCTGGAATTTTAAACAAACGACACAGTGCATTTGTGTCTCTCCACAGACTA 1208
QY 181 CysThrAspAsnGlyIleMetIleAlaTProAsnGlyIleGluArgLeuAlaGlyLeu 200
DB 1209 TGCACGTGATTAATGCAATTAATGATGCAATGATTAAGAAACATCAACGCTGCGCTTG 1268
QY 201 GlyIleLeuHisAspIleGluGlyIleArgTyrGluProLysCysProLeuGlyValAsp 220
DB 1269 GGCATTTTACATACATACATGAAGCATCCGCTATGACCAAAATGTCTCTTGAGTGAAC 1328
QY 221 ILeSerLysGluValGlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 239
DB 1329 ATATCAAAAGACGTTGAGAAAGCTTCATTAAGTACCAACATTTAAATGAGAGATA 1385
RESULT 6
AAD46856
ID AAD46856 standard; cDNA; 1820 BP.
AC AAD46856;
XX 27-JAN-2003 (first entry)
XX Human glycoprotease 28472 cDNA.
XX
XX Human; adenosine deaminase; seven transmembrane domain receptor; cancer;
XX 7TM; glycoprotease; immune disorder; IGA deficiency; allergy; arrhythmia;
XX rheumatoid arthritis; diabetes; atherosclerosis; cardiovascular disorder;
XX hypercenaemia; ischaemic heart disease; obesity; myocardial infarction;
XX endothelial cell disorder; Grave's disease; psoriasis; brain disorder;
XX Parkinson's disease; Alzheimer's disease; haematopoietic disorder;
XX cerebral oedema; metabolic disorder; liver disorder; platelet disorder;
XX chromosome mapping; tissue typing; gene therapy; neuroprotective;
XX cytosolic; anorectic; cardiac; haemostatic; gene; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 146..1390
XX /*tag= a
XX /product= "Human 28472 protein"
XX /note= "This region is specifically claimed as SEQ ID NO:
XX 6 in claim 1 of the specification"
XX
XX WO200274960-A2.
XX
XX 26-SEP-2002.
XX
XX 08-NOV-2001; 2001WO-US051427.
XX
XX 08-NOV-2000; 2000US-0246768P.
XX 08-NOV-2000; 2000US-0246772P.
XX 15-NOV-2000; 2000US-0249185P.
XX
XX (MILL-) MILLENNIUM PHARM INC.
XX
XX Leiby KR, Kapeller-Libermann R, Glucksmann M;
XX
XX WPI; 2002-759898/82.
XX P-FSDB; AAE29234.
XX
XX New 38650, 28472, 5495, 65507, 81588 or 14354 nucleic acid molecules,
XX useful for diagnosing and treating cancer, immune, cardiovascular,
XX hematopoietic, brain, pain, metabolic, liver or platelet disorders, and
XX in pharmacogenomics.
XX
XX Claim 1; Fig 8; 178pp; English.
XX
XX The present invention relates to novel 38650, 28472, 5495, 65507, 81588
XX CC or 14354 nucleic acid molecules which encode adenosine deaminase, glyco-
XX CC protease or seven transmembrane domain (7TM) receptor family members.
XX
XX Sequences of the invention are useful in diagnosing and treating cancer

CC or aberrant cellular proliferation and/or differentiation (e.g. colon or
CC lung cancer), immune disorders (e.g. selective IGA deficiency, rheumatoid
CC arthritis, diabetes, allergy), heart or cardiovascular disorders (e.g.,
CC hyperension, atherosclerosis, arrhythmias, ischaemic heart disease,
CC myocardial infarction, thrombus) including endothelial cell disorders
CC (e.g. psoriasis, Grave's disease), haematopoietic disorders, brain
CC disorders (e.g. cerebral oedema, Parkinson's or Alzheimer's disease),
CC pain and metabolic disorders (e.g. obesity), liver disorders or platelet
CC disorders. They are also useful in screening assays, predictive medicine
CC (e.g. diagnostic assays, prognostic assays, monitoring clinical trials
CC and pharmacogenetics) and prophylactic and therapeutic methods. The
CC nucleic acids may also be used in chromosome mapping, tissue typing and
CC forensic biology and as surrogate markers. Sequences of the invention are
CC also used in gene therapy. The present sequence is human glycoprotease
XX 28472 cDNA
XX
SQ Sequence 1820 BP; 543 A; 365 C; 393 G; 518 T; 0 U; 1 Other;
Alignment Scores:
Pred. No.: 8.22e-128 Length: 1820
Score: 1203.00 Matches: 232
Percent Similarity: 98.33% Conservative: 3
Best Local Similarity: 97.07% Mismatches: 4
Query Match: 97.02% Indels: 0
DB: Gaps: 0
US-10-649-273-2_copy_176_414 (1-239) x AAD46856 (1-1820)
QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20
DB 671 CTGTTGGCATTAATGTTCAAGGAGGATTTTCATTTTCCTTGGAAGCTTTTGACATTA 730
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40
DB 731 GCACCAAGGTGACATGCTTGACAGGTGCAAGAAACATCTTTTAATAACATCCAGAG 790
QY 41 CysSerThrMetSerGlyGlyLysAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPhe 60
DB 791 TGCTCCACCATGAGTGGTGGGAAAGCCATGACATTTGGCCAAACGAAATGATTT 850
QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnGlyAspPheSerPheThrGly 80
DB 851 CATTTTGACATCAAAACCTCCCTGCATCATCTAAATTTGATTTTCTTTTACTGGA 910
QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGlyIleGluLys 100
DB 911 CTTCAACAGCTTACTGATTAATAATTAAGAAACAGAAAGATTTTGAAG 970
QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 120
DB 971 GGGCAAACTCTGCTTTCAGCAGACATTTGCTGCCAAGTACACACACATGCAAT 1030
QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140
DB 1031 CATCTTGGAAGAAACACATCGGGCTATTGCTGTTTGTAAAGCAGAGACTTGTACCT 1090
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaIleAsnPheTyrIleArgArg 160
DB 1091 CAAATAATATGCAATGATGCTGTCATCTGTGTGTCGAAAGTATCTTAATCCGCA 1150
QY 161 AlAlenGluIleLeuThraAsnAlaThrGlnCysThrIleuLeuCysProProProArgLeu 180
DB 1151 GCTCTGGAATTTTAAACAAACGACACAGTGCATTTGTGTCTCTCCACAGACTA 1210
QY 181 CysThrAspAsnGlyIleMetIleAlaTProAsnGlyIleGluArgLeuAlaGlyLeu 200
DB 1211 TGCACGTGATTAATGCAATTAATGATGCAATGATTAAGAAACATCAACGCTGCGCTTG 1270
QY 201 GlyIleLeuHisAspIleGluGlyIleArgTyrGluProLysCysProLeuGlyValAsp 220
DB 1271 GGCATTTTACATACATACATGAAGCATCCGCTATGACCAAAATGTCTCTTGAGTGAAC 1330
QY 221 ILeSerLysGluValGlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 239

DB 1331 ATATCAAAAGAGTTGGAGAGCTTCATTAATAAGTACCACTTAATAATGAGATA 1387

RESULT 7
ACAG0887
ID ACAG0887 standard; cDNA; 1820 BP.
XX
AC AGAG0887;
XX
08-JUL-2003 (first entry)
XX
DE Human cDNA 28472 encoding a glycoprotease.
XX
KW Human; sex; gene; cancer; aberrant cellular proliferation;
KW differentiation; immune disorders; heart disorder; brain disorder;
KW cardiovascular disorder; endothelial cell disorder; pain disorder;
KW haematopoietic disorder; blood vessel disorder; metabolic disorder;
KW liver disorder; platelet disorder; glycoprotease.
XX
OS Homo sapiens.
XX
FH Key location/Qualifiers
FT CDS 146..1390
FT /tag=a
FT /product="Glycoprotease"
FT /note="This CDS is specifically claimed in claim 1"
XX
US2003009017-A1.
XX
09-JAN-2003.
XX
08-NOV-2001; 2001US-00012140.
XX
08-NOV-2000; 2000US-0246768P.
PR 08-NOV-2000; 2000US-0246772P.
PR 15-NOV-2000; 2000US-0249185P.
XX
PA (LEIB/) LEIBY K R.
PA (KAPE/) KAPELLEK-LIBERMANN R.
PA (GLUC/) GLUCKSMANN M A.
XX
PI Leiby KR, Kapellek-Libermann R, Glucksmann MA;
XX
WI; 2003-428888/40.
DR P-SDB; AB09559.
XX
XX
PT New isolated 38650, 28472, 5495, 65507, 81588 and 14354 nucleic acid
PT molecules, useful for diagnosing, treating cancer, pain, or immune,
PT heart, endothelial cell, haematopoietic, blood vessel, brain, metabolic
PT and liver disorders.
XX
PS Claim 2; Fig 8; 90pp; English.
XX
XX The invention relates to an isolated 38650 (encoding adenosine
CC deaminase), 28472 (encoding a glycoprotease), 5495 (encoding a 7
CC transmembrane receptor), 65507, 81588 or 14354 nucleic acid molecule (or
CC a sequence which is at least 60% identical to the six nucleic acids or
CC their open reading frames, fragments of at least 15 nucleotides,
CC naturally occurring variants, or a DNA insert of the plasmid deposited
CC with the American Type Culture Collection as Accession No. not defined in
CC the specification, which encodes the amino acid sequence). Also included
CC are a host cell containing the nucleic acid (used to produce the
CC proteins), the encoded proteins, an antibody that selectively binds to
CC the polypeptide, and identifying a compound that binds to/modulates the
CC activity of the polypeptide. The nucleic acid molecules, polypeptides and
CC methods are useful for diagnosing, treating cancer, aberrant cellular
CC proliferation and/or differentiation, immune disorders, heart disorders,
CC cardiovascular disorders including endothelial cell disorders,
CC haematopoietic disorders, blood vessel disorders, brain disorders, pain
CC and metabolic disorders, liver disorders and platelet disorders (many
CC examples of these disorders are given in the specification). The present
CC sequence is the Human cDNA 28472 encoding a glycoprotease

Sequence 1820 BP; 543 A; 365 C; 393 G; 518 T; 0 U; 1 Other:

Alignment Scores:

Pred. No.:	8, 22e-128	Length:	1820
Score:	1203.00	Matches:	232
Percent Similarity:	98.33%	Conservative:	3
Best Local Similarity:	97.07%	Mismatches:	4
Query Match:	97.02%	Indels:	0
DB:	8	Gaps:	0

US-10-649-273-2_COPY_176_414 (1-239) x ACAG0887 (1-1820)

QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20
DB 671 CTTGTTGGCATTTAGTTCAAGAGTTTCAAGATTTCTCTTCTTGGAAAGCTTTGGACATA 730
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuLysHisProGlu 40
DB 731 GCACCAAGTGACATGTTGACAAAGGTGGCAAGAGACTTTCTTAATAAACATCCACAG 790
QY 41 CysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60
DB 791 TCTCCACCATAGAGTGAGTGGAAAGCCATAGAACATTTGGCCAAACAGAAATAGATT 850
QY 61 HisPheAspIleLysProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80
DB 851 CATTTTGCATCAAACTCCCTTGATCATGCTAAATAATGTCATTTCTTTTACTGGA 910
QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGlnLysGlnGlyLysIleGlyLys 100
DB 911 CTTCAACAGCTTACGATTAATAATATGAAACAGGAAACAGAGAGATTTGGAAG 970
QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleValGlnHisThrMetAlaCys 120
DB 971 GGGCAATCTGTCTTCACAGACAGCATTTGCTCCACAGTACAGACCAATGCGATGT 1030
QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140
DB 1031 CATCTTGGAAAGAACATCATGGGCTATTCTTTTGTAGACAGACATTTGTAACCT 1090
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheThrIleArgArg 160
DB 1091 CAAATTAATGCACTGCTGCTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1150
QY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProAlaGlu 180
DB 1151 GCTCTGAAATTTTAACAAAGCAACAGTGCATTTGTGTCTCTCCACACATCA 1210
QY 181 CysThrAspAsnGlyIleMetIleAlaTPAsnGlyIleGluArgLeuArgAlaGlyLeu 200
DB 1211 TGCACCTGATTAATGCACTTATGATTTGATGGAATGATTTGGAAGACTGCTGCTG 1270
QY 201 GlyIleLeuHisAspIleGlnGlyIleArgIleArgIleProLysCysProLeuGlyValAsp 220
DB 1271 GGCAATTTTACATGACATAGAAAGCATCCGCTATGAAACAAATATGCTCTTGGAGTAGAC 1330
QY 221 HisLeuValGlnValGlyLysAlaSerIleLysValProGlnLeuLysMetGluIle 239
DB 1331 ATATCAAAAGAGTTGGAGAGCTTCATTAATAAGTACCACTTAATAATGAGATA 1387

RESULT 8
ID ABS57020
ID ABS57020 standard; cDNA; 1821 BP.
XX
AC ABS57020;
XX
XX 30-JAN-2003 (first entry)
XX
DE cDNA encoding novel human glycoprotease 28472.
XX
KW Cancer; aberrant cell proliferation; aberrant cell differentiation;
KW breast cancer; ovarian cancer; prostate cancer; colon cancer;
KW lung cancer; immune disorder; heart disorder; cardiovascular disorder;

KW endothelial disorder; hematopoietic disorder; blood vessel disorder;
 KW brain disorder; pain; metabolic disorder; liver disorder; diabetes;
 KW platelet disorder; carcinoma; sarcoma; leukemia; Hodgkin's disease;
 KW autoimmune disorder; hypertension; atherosclerosis; heart failure;
 KW myocardial infarction; ischemic heart disease; Crohn's disease;
 KW Grave's disease; Kawasaki syndrome; Raynaud's disease; aneurysm;
 KW cerebral ischemia; peripheral neuropathy; Alzheimer's disease;
 KW Parkinson's disease; anorexia nervosa; cachexia; glycoprotease; gene; ss.
 XX
 OS Homo sapiens.
 FH
 FT Key Location/Qualifiers
 FT CDS 147..1391
 FT /tag= a
 FT /product= "Glycoprotease 28472"
 FT /note= "Specifically claimed in claim 1"
 FT
 PN WO200277233-A2.
 XX
 PD 03-OCT-2002.
 XX
 PF 08-NOV-2001; 2001WO-US046724.
 XX
 PR 08-NOV-2000; 2000US-0246768P.
 PR 08-NOV-2000; 2000US-0246772P.
 PR 15-NOV-2000; 2000US-0249185P.
 XX
 PA (MILL.) MILLENNIUM PHARM INC.
 PI Leiby KR, Kapeller-Libermann R, Glucksmann M;
 DR WPI: 2003-029938/02.
 DR P-PSDB; ABG71162.
 XX
 PT New adenosine deaminase, glycoprotease and seven transmembrane domain
 PT nucleic acids and polypeptides, designated 38650, 28472, 5495, 65507,
 PT 81588 and 14354, useful for treating e.g. leukemias, Hodgkin's disease or
 PT hypertension.
 XX
 PS Claim 2; Fig 8A-B; 178pp; English.
 XX
 CC The invention describes isolated 38650, 28472, 5495, 65507, 81588 and
 CC 14354 nucleic acid molecules (I) and their encoded polypeptides (II). The
 CC 38650 nucleic acid molecule comprises a sequence encoding adenosine
 CC deaminase. The 28472 nucleic acid molecule comprises a sequence encoding
 CC a human glycoprotease. The 5495, 65507, 81588 and 14354 comprise
 CC sequences that encode a human seven transmembrane domain (7TM). The
 CC 38650, 28472, 5495, 65507, 81588 or 14354 nucleic acid and polypeptide
 CC sequences are useful for diagnosing, preventing or treating a subject
 CC with or at risk of developing a disorder, e.g. cancer or aberrant
 CC cellular proliferation and/or differentiation (e.g. breast, ovarian,
 CC prostate, colon or lung cancer), immune disorders, heart disorders,
 CC cardiovascular disorders, endothelial disorders, hematopoietic disorders,
 CC blood vessel disorders, brain disorders, pain and metabolic disorders,
 CC liver disorders or platelet disorders. These disorders include carcinoma,
 CC sarcoma, leukemia, Hodgkin's disease, autoimmune disorders,
 CC hypertension, atherosclerosis, heart failure, myocardial infarction,
 CC ischaemic heart disease, Crohn's disease, Grave's disease, Kawasaki
 CC syndrome, Raynaud's disease, aneurysm, cerebral ischemia, peripheral
 CC neuropathy, Alzheimer's disease, Parkinson's disease, anorexia nervosa,
 CC cachexia or diabetes. This sequence encodes the novel human glycoprotease
 CC 28472
 XX
 SQ Sequence 1821 BP, 543 A; 365 C; 394 G; 518 T; 0 U; 1 Other;

US-10-649-273-2_COPY_176_414 (1-239) x ABS57020 (1-1821)
 QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20
 DB 672 CTTGGGCACTTACGTTCAAGAGGTTTCAAGTTTCTGCTTGGAAAGCTTTGGACATA 731
 QY 21 AlaProGlyAspMetLeuAspIleValAlaArgArgLeuSerLeuIleLysHisProGlu 40
 DB 732 GCACCAAGGTCATCTTGCACCAAGGTGCAAGAGCTTTCTTATATAAATCCACAG 791
 QY 41 CysSerThrMetSerGlyGlyValSerAlaIleGlnHisLeuAlaGlnGlyValAsnArgPhe 60
 DB 792 TGTCTCACCACGATGAGTGGGGAAGCCATAGAACATTTGGCCAAACAGAGAAATAGATTT 851
 QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80
 DB 852 CATTTTGACATCAAACTCCCTTGCATCATGCTAAATAATTTGATTTTCTTTATCTGGA 911
 QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGlnLysGlnGlyIleGluLys 100
 DB 912 CTTCAACACGTTACTGATTAATAATATGAAACAGGAACAGAGAGATTTGAGAAG 971
 QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCys 120
 DB 972 GGGCAAAATCTGCTTTCAGCAGCAGCATTTGCTGCCACAGTACAGCACAATGGCATGT 1031
 QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140
 DB 1032 CATCTTGGAAAGAACACATCGGGCTATTCTGTTTGTAGCAGAGACTTTGTACT 1091
 QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheTyrIleArgArg 160
 DB 1092 CAAATTAATGACGTACTGTTGCATCTGTTGCTGCAAGTACTTATATCCACAG 1151
 QY 161 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProPheArgLeu 180
 DB 1152 GCTTGGAAATTTTAAACCAACCAACAGTGCATCTTGTGTCTCTCCCAACATA 1211
 QY 181 CysThrAspAsnGlyIleMetIleAlaTyrAsnGlyIleGluArgLeuArgAlaGlyLeu 200
 DB 1212 TGCACGTGATTAATGCGATTAATGATGATGATGATGATGATGATGATGATGATGATG 1271
 QY 201 GlyIleLeuHisAspIleGlnGlyIleArgTyrGluProLysCysAspProLeuGlyValAsp 220
 DB 1272 GGCATTTTACATGACATTAAGGCAATCCGCTATGAAACCAAAATGCTCTTGGAGTAGAC 1331
 QY 221 HisLeuGluValGlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 239
 DB 1332 ATATCAAAAGAAAGTTGGAGAGCTTCCATTAAGTACCAATTAATAAATGAGAGATA 1388
 DB
 RESULT 9
 ADS52832
 ID ADS52832 standard; cDNA; 2208 BP.
 XX
 AC ADS52832;
 XX
 DT 20-NOV-2003 (first entry)
 XX
 DE Human coding sequence, SEQ ID 400.
 XX
 XX Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;
 KW Gene Therapy; human; secretory protein; membrane proteins; cancer;
 KW inflammatory disease; osteoporosis; neurological disease; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN EP1293569-A2.
 XX
 PD 19-MAR-2003.
 XX
 PP 21-MAR-2002; 2002EP-00006586.
 XX
 PR 14-SEP-2001; 2001JP-00328381.

PR 24-JAN-2002; 2002US-0350435P.
XX (HELI-) HELIX RES INST.
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI Isegami T, Sugiyama T, Otsuki T, Makamatsu A, Sato H, Ishii S;
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuno Y;
XX WPI; 2003-395539/38.
DR P-PSDB; ADA54471.
XX
PT New polynucleotides encoding full-length polypeptides, e.g. secretory
PT and/or membrane proteins, useful for developing medicines for diseases in
PT which the gene is involved, or as target molecules for gene therapy.
XX
PS Claim 1; SEQ ID NO 400; 205bp; English.
XX
CC The present invention relates to novel human secretory or membrane
CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
CC ADA54071). The coding sequences are useful in the gene therapy of
CC diseases caused by abnormalities of the proteins, e.g. cancer,
CC inflammatory diseases, osteoporosis or neurological disease.
XX
SQ Sequence 2208 BP; 660 A; 454 C; 451 G; 643 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 3,88e-111 Length: 2208
Score: 1059.00 Matches: 211
Percent Similarity: 88.28% Conservative: 0
Best Local Similarity: 88.28% Mismatches: 4
Query Match: 85.40% Indels: 24
DB: 10 Gaps: 1

US-10-649-273-2_COPY_176_414 (1-239) x ADA52832 (1-2208)

QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20
DB CTGTTGGCATTAGTTCAGAGAGTTTCAGATTTTCGCTTCTTGGAAGCTTTGGACATA 928

QY 21 AlaProGlyAspMetLeuAspLysValAlaArgAlaGluSerLeuIleLysHisProGlu 40
DB GCACACAGGTGACATGCTTGACACAGGTGACAGAACATCTCTTAATAAATCAGAG 988

QY 41 CysSerThrMetSerGlyGlyLysAlaIleGlnHisLeuAlaValGlnGlyAsnArgPhe 60
DB TGCTCCACCATGAGTGGTGGGAAACCATAGAACATTTGGCCAAACAGAAATGATTT 1048

QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80
DB CATTTTGACATCAAACTCCCTTGACATCATGCTAATAAATTTGATTTCTTTACTGCA 1108

QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGluGlyIleGluLys 100
DB CTTCAACACGTTACTGATTAATAATATGAAAAAGAAAAAGAAAGATTTGAGAAAG 1168

QY 101 GlyGlnIleLeuSerSerSerAlaAlaAspIleAlaIleAlaThrValGlnHisThrMetLacys 120
DB GGGCAAAATCTGTCTTGACAGACACATTCCTGCACAGTACACACACAAATGCAATGT 1228

QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140
DB CATCTTGTGAAGAAGACATCGGCTATTCTGTTTGTGTAACAGAGACTGTGTAACCT 1288

QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyValAlaAsnAsnPheTyrIleArgArg 160
DB CAAATATATGACAGTACTGTTGACATCTGTGTGTGCGAAGTAACTTCTGTATCGGACA 1348

QY 161 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeu 180
DB GCTCTGGAAATTTTAAACAAACGCAACAGTGCATCTTGTGTGTCTCTCCAGACTA 1408

QY 181 CysThrAspAsnGlyIleMetIleAlaThrAsnGlyIleGluArgLeuArgAlaGlyLeu 200

DB 1409 TGCATCTGATTAATGCAATTAGATTGCA----- 1435
QY 201 GlyIleLeuHisAspIleGluGlyIleArgTyrGluProLysCysProLeuGlyValAsp 220
DB 1436 -----TGAATGCTCTTGGAGATGAC 1456
QY 221 HisSerLysGluValAlaGlyAlaSerIleLysValProGlnLeuLysMetGluIle 239
DB 1457 ATATCAAAAGAAAGTGAAGAAAGCTTCATPAAAGTACCAATTTAAAAATGAGATA 1513

RESULT 10
ADQ24627
ID ADQ24627 standard; DNA; 2890 BP.
XX
AC ADQ24627;
XX
DT 26-AUG-2004 (first entry)
XX
DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 7447.
XX
KW soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
KW ds.
XX
OS Homo sapiens.
XX
PN WO2004048938-A2.
XX
PD 10-JUN-2004.
XX
PF 26-NOV-2003; 2003WO-US038193.
XX
PR 26-NOV-2002; 2002US-0429739P.
XX
PA (PROT-) PROTEIN DESIGN LABS INC.
XX
PI Aziz N, Ginsburg WM, Zlotnik A;
XX
DR WPI; 2004-441208/41.
XX
PT Early detection of soft tissue sarcoma comprises determining expression
PT of a gene in a first soft tissue sample and a normal soft tissue sample
PT and comparing the gene expression, also useful in treating soft tissue
PT sarcoma.
XX
PS Example 2; SEQ ID NO 7447; 210bp; English.
XX
CC The invention relates to a novel method for detecting soft tissue sarcoma
CC which comprises obtaining a first soft tissue sample from an individual
CC and a normal soft tissue sample from the same or different individual,
CC determining the expression of a gene in both samples and comparing the
CC expression of the gene in both soft tissue samples, where a higher level
CC of protein expression in the first soft tissue sample indicates the
CC presence of soft tissue sarcoma. The method of the invention has
CC cytostatic applications and may be useful for detecting soft tissue
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC acid sequences may be useful in diagnostic and screening applications.
CC The current sequence is that of a human soft tissue sarcoma-upregulated
CC DNA of the invention. The current sequence is not shown within the
CC specification per se but was submitted in CD format by the inventor.
XX
SQ Sequence 2890 BP; 869 A; 609 C; 611 G; 789 T; 0 U; 12 Other;

Alignment Scores:
Pred. No.: 5.67e-111 Length: 2890
Score: 1059.00 Matches: 211
Percent Similarity: 88.28% Conservative: 0
Best Local Similarity: 88.28% Mismatches: 4
Query Match: 85.40% Indels: 24
DB: 12 Gaps: 1

US-10-649-273-2_COPY_176_414 (1-239) x ADQ24627 (1-2890)

QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20
 DB 1526 CTGTTGGCATTAAGTTCAAGAGAGTTTCAAGATTTTCTGCTTCTGGAAGCTTTGACATA 1585
 QY 21 AlaProGlyAspMetLeuAspLysValAlaArgAlaGluSerLeuIleLysHisProGlu 40
 DB 1586 GCACACAGTGAACGCTTGAACAGGTCAGAACACATCTCTTATATATAACATCCAGG 1645
 QY 41 CysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60
 DB 1646 TGTCTCCACATGAGTGGTGGGAAAGCCATAGAACCTTTGGCCAAACAGAAATATGATTT 1705
 QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCyAspPheSerPheThrGly 80
 DB 1706 CATTTTGACATCAACCTCCCTTGACATCATGCTAAATATGATTTTCTTTTCTGGA 1765
 QY 81 LeuGlnHisValIleThrAspLysIleIleMetLysLysGlnGlyGluGlnGlyIleGluLys 100
 DB 1766 CTTCACACGCTTACTGATTAATAATTAATGAAAAAGAAAAAGAGAAAGTATTGAGAG 1825
 QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 120
 DB 1826 GGGGAAATCTGTCTTTCAGACGACAGCATGCTGCACAGTACAGACACAAATGGCATGT 1885
 QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140
 DB 1886 CATTTTGGAAAGAACACATCGGGCTATTTCTGTTTGTAAACAGAGACCTGTTTACT 1945
 QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheTyrIleArgArg 160
 DB 1946 CAAATATATGACGCTGCTGTCATCTGTCGTCGCAAGTACCTTGTATCCGACAG 2005
 QY 161 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeu 180
 DB 2006 GCTCTGAAATTTTAAACAAACGCAACACGTCACCTTGTGTGTCTCTCCACAGACTA 2065
 QY 181 CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeu 200
 DB 2066 TGCACTGATTAATGGCATTAATGATTCGA----- 2092
 QY 201 GlyIleLeuHisAspIleGlnGlyIleArgTyrGluProLysCysProLeuGlnValAlaAsp 220
 DB 2093 -----TGATGTCTCTTGAGAGTAGAC 2113
 QY 221 IleSerLysGluValGlyGlnAlaSerIleLysValProGlnLeuLysMetGluIle 239
 DB 2114 ATATCAAAAGAAAGTTGAGAAAGCTTCCATTAAGTACCAAAATTAATAATGAGATA 2170
 RESULT 11
 ADE31345/c
 ID ADE31345 standard; DNA, 3358 BP.
 XX
 AC ADE31345;
 XX
 DT 29-JAN-2004 (first entry)
 XX
 DE Human diagnostic and therapeutic polynucleotide (dithp), SEQ ID NO 100.
 XX
 KW diagnostic and therapeutic polynucleotide; dithp; antiarteriosclerotic;
 KW antiinflammatory; cerebroprotective; antidiabetic; antidiabetic;
 KW immunosuppressive; neuroprotective; nootropic; neuroleptic; tranquillizer;
 KW osteopathic; antirheumatic; antineumatic; cytostatic; hepatotropic;
 KW vitaminic; haemostatic; anti-HIV; antithyroid; thyromimetic;
 KW dermatological; antibacterial; fungicide; antiparasitic; anticonvulsant;
 KW chromolytic; anticoagulant; anorectic; vasorectic; antidiabetic;
 KW gene therapy; protein replacement therapy; human; gene; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO2003062376-A2.
 XX
 PD 31-JUL-2003.
 XX

PF 13-JAN-2003; 2003WO-US001096.
 XX
 PR 16-JAN-2002; 2002US-0349384P.
 PR 17-JAN-2002; 2002US-0349413P.
 PR 17-JAN-2002; 2002US-0349946P.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 XX
 PI Jones AL, Dahl CR, Gietzen D, Chinn J, Dufour GE, Jackson JI,
 PI Yu JY, Tuson O, Yap PE, Amshay SR, Dam TC, Liu TF, Gerslin EH;
 PI Peralta CH, Lewis SA, Chen A, Marwaha R, Ian RV, Urashka ME;
 PI Kristnam SR, Kolluru V, Panesar IS;
 DR WPI; 2003-636732/60.
 DR P-PSDB; ADE31156.
 XX
 PT New human diagnostic and therapeutic polynucleotides and polypeptides,
 PT useful for diagnosing, treating or preventing e.g. leukemia, brain
 PT cancer, atherosclerosis, AIDS, thyroiditis, infections, obesity, stroke
 PT or Alzheimer's.
 PS
 XX Claim 1: SEQ ID NO 100; 634pp; English.
 XX
 CC The invention relates to a novel isolated human diagnostic and
 CC therapeutic polynucleotide (designated dithp). The novel dithp
 CC polynucleotide comprises: any of 188 DNA sequences consisting of 195-7798
 CC base pairs fully defined in the specification; a polynucleotide
 CC comprising a naturally occurring polynucleotide sequence at least 90%
 CC identical to the dithp polynucleotide; a polynucleotide complementary to
 CC the dithp polynucleotide or its polynucleotide which is at least 90%
 CC identical; or an RNA equivalent of any of the polynucleotides mentioned
 CC above. The dithp polynucleotides have the following activities:
 CC antiarteriosclerotic, antiinflammatory, cerebroprotective, antidiabetic,
 CC antidiabetic, immunosuppressive, neuroprotective, nootropic, neuroleptic,
 CC tranquillizer, osteopathic, antirheumatic, antineumatic, cytostatic,
 CC hepatocytic, vitaminic, haemostatic, anti-HIV, antithyroid, thyromimetic,
 CC dermatological, antibacterial, fungicide, antiparasitic, anticonvulsant,
 CC thrombolytic, anticoagulant, anorectic, vasorectic, and antidiabetic. The
 CC novel dithp polynucleotides polypeptide can be used in gene therapy and
 CC protein replacement therapy. The dithp polynucleotides or dithp
 CC polypeptides are useful for diagnosing, preventing or treating diseases
 CC associated with the expression of human molecules. In particular, these
 CC diseases include cancers (e.g. adenocarcinoma, leukemia, melanoma, brain
 CC cancer, breast cancer, cervix cancer, bone cancer, liver cancer, lung
 CC cancer) or other cell proliferative disorders (e.g. arteriosclerosis,
 CC atherosclerosis, cirrhosis, hepatitis, polycythemia vera, primary
 CC chromocytopenia), autoimmune/inflammatory disorders (e.g. AIDS,
 CC Addison's disease, thyroiditis, Crohn's disease, Graves' disease,
 CC Hashimoto's thyroiditis, multiple sclerosis, osteoarthritis, rheumatoid
 CC arthritis, scleroderma, systemic lupus erythematosus), infections (e.g.
 CC viral, bacterial, fungal or parasitic infection), developmental disorders
 CC (e.g. Cushing's syndrome or epilepsy), endocrine disorders (e.g.
 CC chromoblasts, hypopituitarism, hypogonadism, gigantism, goiter) metabolic
 CC disorders (e.g. hypercholesterolaemia, hypoglycaemia, diabetes,
 CC hyperlipidaemia, obesity), neurological disorders (e.g. ischaemic
 CC cerebrovascular disease, stroke, Alzheimer's disease, Pick's disease,
 CC Huntington's disease, Parkinson's disease, Creutzfeldt-Jakob disease,
 CC anxiety, schizophrenia), gastrointestinal disorders (e.g. ulcers),
 CC transport disorders (e.g. akinesia or multidrug resistance), or
 CC connective tissue disorders (e.g. Paget's disease or rickets). This
 CC polynucleotide sequence represents one of the human dithp DNA sequences
 CC of the invention.
 XX
 SQ Sequence 3358 BP; 1105 A; 577 C; 601 G; 1075 T; 0 U; 0 Other;
 XX
 Alignment Scores:
 Pred. No.: 1,4e-103 Length: 3358
 Score: 995.50 Matches: 209
 Percent Similarity: 87.82% Conservative: 0
 Best Local Similarity: 87.82% Mismatches: 2
 Query Match: 80.28% Indels: 29
 DB: 10 Gaps: 1

```

US-10-649-273-2_copy_176_414 (1-239) x ADB31345 (1-3358)
QY 29 ValAlaArgArgLeuSerLeuIleLysHisProGluCysSerThrMetSerGly-GlyLeu 48
Db 2104 GTGGCAAGAAAGACTTTCTTATATAAACATCCAGAGTGCTCCACCATGATGCTGTGGGAA 2045
QY 48 GAlAlIleGluHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProle 68
Db 2044 AGCCATAGAAACATTTGGCCAAACAAAGAAATAGATTTCATTTTGACATCAAACTCCCTT 1985
QY 68 uHisHAlaLysAsnGlyAspPheSerPheThrGlyLeuGlnHisValThrAspLysIle 88
Db 1984 GCATCATGCTTAAATATGTAATTTCTTTTCTTATGAGCTTCAACAGCTTACTGATTAAT 1925
QY 88 eIleMetLysLysGluLysGluLys-----GlyIleGluLysGlyGlnI 96
Db 1924 AATATATGAAAAAGAAAAAGAAAGATATATTTCTAATTAAGTAAAGTGAACAGATAA 1865
QY 97 -----GlyIleGluLysGlyGlnI 103
Db 1864 TATTCTGATTTGCTCTAAATAATAGCTCTCATTTCTGACAGTATTGGAAGGGGCAAA 1805
QY 103 IeLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCysHisLeuVal 123
Db 1804 TCTGTCTTCAGCAGCAGACATTCCTGCCACAGTACACACACATGGCATGTCATCTTG 1745
QY 123 alLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuProGlnAsnA 143
Db 1744 TGAAAAAGAACACATCGGCTATTTCTGTTTGTAAACAGAGACTTTTACTCTCAAAATA 1685
QY 143 snAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheTyrIleArgArgAlaLeuG 163
Db 1684 ATGAGACTGCTGTGATCTGT-GGTGTGCGAAGTACTTCTATATCCGACAGCTCTGG 1626
QY 163 IuIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeuCysThrA 183
Db 1625 AATATTTTAAAC-AAGCAACACAGAGTACCTTTGTGTGCTCTCCAGCATATCACAG 1567
QY 183 sPAngIyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIle 203
Db 1566 ATAATGGCATTAATGATTCATGAGAAATGTAATTAAGACTACGCTGCTGGCATTT 1507
QY 203 euHisAspIleGluGlyIleArgTyrGluProLysCysProLeuGlyValAspIleSer 223
Db 1506 TACTGATCATGAGAGGATCCGCTATGAAACCAAAATGCTCTTGAAGTAAAGCATATCA 1447
QY 223 ysgIuValGlyLysValSerIleLysValProGlnLeuLysMetGluIle 239
Db 1446 AAGAAGTTGGAAGAGCTTCCATAAAGTAAACATTAATAATGAGAGATA 1397
RESULT 12
AB075508 standard; DNA; 1572 BP.
XX AB075508;
XX
XX
XX 07-NOV-2002 (first entry)
XX
XX Murine sialoglycoprotease-like gene sequence SEQ ID NO:7.
XX
XX Murine; mouse; protease; calcium activated neutral protease type 5;
XX CAPN5; trypsin 4; sialoglycoprotease; enzyme; genetic disease;
XX neurological; neuropsychological; psychotic illness; transgenic animal;
XX gene; ds.
XX
XX Mus musculus.
XX
XX PN M0200245491-A2.
XX
XX 13-JUN-2002.
XX
XX 05-DEC-2001; 2001WO-US046405.
XX

```

```

PR 06-DEC-2000; 2000US-0251803P.
PR 06-DEC-2000; 2000US-0251820P.
PR 13-DEC-2000; 2000US-0255971P.
PA (DELT-) DELTAGEN INC.
XX
XX Allen KD, Leviten MW;
XX
XX WPI; 2002-657389/70.
XX
XX Novel transgenic animal, comprising a disruption in protease target gene,
XX PT is useful for identifying agents that ameliorates a phenotype associated
XX with a disruption in a protease target gene.
XX
XX Example 3; Fig 7; 62pp; English.
XX
XX CC The present invention describes a non-human transgenic animal (I)
XX CC comprising a disruption in a protease target gene (PG) selected from
XX CC calcium activated neutral protease type 5 (CAPN5) gene, trypsin 4 gene
XX CC and sialoglycoprotease-like gene. Also described is a targeting construct
XX CC (II), comprising a first polynucleotide sequence homologous to at least a
XX CC first portion of PG, a second polynucleotide sequence homologous to at
XX CC least a second portion of PG and a selectable marker. (II) is useful for
XX CC producing a transgenic mouse comprising a disruption in a protease target
XX CC gene, by introducing (II) into a cell, introducing the cell into a
XX CC blastocyst, implanting the resulting blastocyst into a pseudopregnant
XX CC mouse, where the pseudopregnant mouse gives birth to a chimeric mouse,
XX CC and breeding the chimeric mouse to produce the transgenic mouse. (I) is
XX CC useful for identifying an agent that modulates the expression or function
XX CC of a protease target gene, by administering an agent to (I) and
XX CC determining whether the expression or function of the disrupted protease
XX CC target gene in (I) is modulated. (I) is also useful for testing the
XX CC efficacy of proposed genetic and pharmacological therapies for human
XX CC genetic diseases, such as neurological, neuropsychological or psychotic
XX CC illnesses. The present sequence represents murine sialoglycoprotease-like
XX CC gene sequence, which is used in an example from the present invention
XX
XX SQ Sequence 1572 BP; 459 A; 337 C; 340 G; 429 T; 0 U; 7 Other;
XX
XX Alignment Scores:
XX Pred. No.: 1.28e-89 Length: 1572
XX Score: 870.00 Matches: 184
XX Percent Similarity: 82.77% Conservative: 13
XX Best Local Similarity: 77.31% Mismatches: 37
XX Query Match: 70.16% Indels: 5
XX DB: 6 Gaps: 2
XX
XX US-10-649-273-2_COPY_176_414 (1-239) x AB075508 (1-1572)
QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlnGlyLysSerLeuAspIle 20
Db 528 CTGTTGCKRTTATGTCACAGAGTGTTCCGATTCCTCTCTGGAAGGCTTTTGACATA 587
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40
Db 588 GCACACAGGCGCATGCTTGAACAAGTGGCAAGAAGCTTTCTTATTAACAACATCCAGAA 647
QY 41 CysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60
Db 648 TGTTCATCAATGAGTGTGGAAGAGCTATAGAACAGTTGCGCCAAACCGAAATATGATTC 707
QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnGlyAspPheSerPheThrGly 80
Db 708 CATTTTACTATCAATTCACCTATGCAATGCAATGCTTAACAATTTGTCATTTTCTTACGCGGA 767
QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGlnGlyIleGluLys 100
Db 768 CTTCAACATTAATTAAGTAAAGTAAACACACAGAAAAAGAAAGGCAATTTGAGAG 827
QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 120
Db 828 GGGCAATCTCTGTCAATCAGCTCAGACATTTGCTGCTGCGTACAGCATCAACAGCGTGC 887

```

QY 121 HisLeuValIysArgThrHisArgAlaIleLeuPheCysIysGlnArgAspLeuLeuPro 140
CC |||||
CC CACCTTCGAAAGAACACATCGCGCTATTCGTTTTCAGACAGAAATAATTTGCTCT 947
Db 888
QY 141 GlnAsnAsnAlaValIleuValAlaSerGlyGlyValAlaSerAsnPhenTyrIleArgArg 160
CC |||||
CC 948 CCAAGCTAACCGAGTATGATGTGATCTCGAGAGTGTGCAGAACTTGTACATCCGAAA 1007
Db 161 AlLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeu 180
QY 1008 GCATTGGAAATTTGCGCAATATGACACCGCATTCGCTTTGTGTCTCACT-TCAGAGCTG 1066
Db 181 CysThrAspAsnGlyIleMetIleAlaThrAsnGlyIleGlnArgLeuArgAlaGlyLeu 200
QY 1067 TGCACCTGACATGCGCATATGATTCATGATGAAATGGAATTAAGTTCGCTGCTT 1126
Db 201 GlyIleLeu-HisAsp-IleGluGlyIleArgTyrGluProIysCysProLeuGlyVal- 219
QY 1127 GGCCTTTACCATCATGATGATAGACATTCGGTATTTAAACCAAAATGTCCTCTTGAGTGA 1186
QY 220 --AspIleSerIysGluValGlyGluAlaSerIleIysValProGlnLeu 235
Db 1187 GGCATTATCCCGAAGAGTTGGCAGA--AGCTTGCCCATTTAAAAAGTTA 1233
RESULT 13
ID AAS84622 standard; cDNA; 2734 BP.
AC AAS84622;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #20426.
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
OS Homo sapiens.
XX
PN W0200175067-A2.
XX
PD 11-OCT-2001.
XX
PE 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Dirmnac RT, Liu C, Tang YT;
XX
PI WPI; 2001-639362/73.
DR P-PSDB; ABG20435.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 1; SEQ ID NO 20426; 103pp; English.

CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 2734 BP; 773 A; 545 C; 639 G; 763 T; 0 U; 14 Other;
Alignment Scores:
Pred. No.: 4.02e-58 Length: 2734
Score: 599.00 Matches: 131
Percent Similarity: 71.57% Conservative: 10
Best Local Similarity: 66.50% Mismatches: 22
Query Match: 48.31% Indels: 34
DB: 5 Gaps: 6
US-10-649-273-2_COPY_176_414 (1-239) x AAS84622 (1-2734)
QY 43 ThrMetSerGlyGlyValAlaIleGluHisLeuAlaIysGlnGlyAsnArgPheHisPhe 62
Db 206 ACCCTGCTAGGGGGGCGCCGCT-----AGATTCCAT----- 235
QY 63 AspIleIysProProLeuHisHisAlaIys-----AsnCys----- 74
Db 236 -----CCTATTTCTCCGATGAAGATACAGGCTGTGTAGCAAGCGTGGACACA 283
QY 75 -----AspPheSerPheThrGlyLeuGlnHisValThrAsp 86
Db 284 TCTGCTTCGGGAGGCCCGACGAGAGCTTTTACTCATCGCGAAGCAAAFTGGAGATAG 343
QY 87 IysIleIleMetIysIysGluIysGluIu-----GlyIleGluIysGlyGlnIleLeu 104
Db 344 CGCTTTACATGCAGAGACGACGACGAGAGAGGGGGGTATTGAGAAAGGGGCAAACTTG 403
QY 105 SerSerAlaAlaAspIleAlaIleAlaThrValGlnHisThrMetAlaCysHisLeuValIys 124
Db 404 TCTTCAGACGACAGACATTCGTCTGCACAGTACGACACATGCGATGCTTGTGAAA 463
QY 125 ArgThrHisArgAlaIleLeuPheCysIysGlnArgAspLeuProGlnAsnAla 144
Db 464 AGAACATCGGGCTATTCTGTTTGTATAGCAGAGACTTGTACTCTCAAAATATATGCA 523
QY 145 ValLeuValAlaSerGlyGlyValAlaAsnAsnPhenTyrIleArgArgAlaLeuGluIle 164
Db 524 GTACTGTTGCACTCTGCTGTGTGCGCAAGTAATTCTATATCCGACAGACTCGAAATT 583
QY 165 LeuThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeuCysThrAsnAsn 184
Db 584 TTAAACAAACGCAACACAGTGCACCTTGTGTGTCTCTCCACGACTATGACATGATAT 643
QY 185 GlyIleMetIleAlaThrAsnGlyIleGlnArgLeuArgAlaGlyLeuGlyIleLeuHis 204
Db 644 GCATTATAAATGATGATGATATGTAAGACTACGTCGCTGGCATTTTATCAT 703
QY 205 AspIleGluGlyIleArgTyrGluProIys-----CysProLeuGlyVal 219
Db 704 GACATAGAAAGCATCCGCTATGAAACCAAGTATGTGGCTCTTCAAGGGCTG 754
RESULT 14
ID ADL86725
ADL86725 standard; DNA; 371 BP.
AC ADL86725;
XX
XX 20-MAY-2004 (first entry)
DT
DE DNA up-regulated in murine multipotent progenitor cells SegID 3118.
XX


```

Db      2 GACATTGCTGCTGCGGTACAGCATGCAACAGCGTGCCACTTGGCAAAAACACATCGC 61
Qy      129 AlaIleuPheCysLysGlnArgAspleuLeuProGlnAsnAsnAlaValIleuValAla 148
      |||||
Db      62 GCTATTCTGTTTTCACAGCAGAGAAATTGCTCTCTCCAGCTAACGCAATATTAGTTGTA 121
      |||||
Qy      149 SerGlyGlyValAlaSerAsnPheTyrIleArgArgAlaLeuGluIleLeuThrAsnAla 168
      |||||
Db      122 TCTGGAGGTGTGCAAGTAACCTTGTAATCCGAAAGCATTGGAATGTCGCAATGCA 181
      |||||
Qy      169 ThrGlnCysThrIleuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIle 188
      |||||
Db      182 ACGCAGTGCAAGTGTGTGTCTCCACTTCCAAAGACTGTGCACATGCACATGCATGAT 241
      |||||
Qy      189 AlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGly 208
      |||||
Db      242 GCATGGAAATGGAAATGGAAATGGAAATGGAAATGGAAATGGAAATGGAAATGGAA 301
      |||||
Qy      209 IleArgTyrGluProLysCysProLeuGlyValAspIleSerLysGluValGlyValAla 228
      |||||
Db      302 ATCCNMTATGACCAAAATNTCNTNTTGTGAGTAGACATNTCCAGAGAAAGTTGCAGAGCT 361
      |||||
Qy      229 SerIleLys 231
      |||||
Db      362 GCCATTAATA 370

```

Search completed: November 10, 2005, 17:34:22
 Job time : 547.789 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: November 10, 2005, 16:53:13 ; Search time 3387.57 Seconds
(without alignment)
2685.516 Million cell updates/sec

Title: US-10-649-273-2_COPY_176_414

Perfect score: 1240
Sequence: 1 L1ALVQGVSDPLLGKSLDI.....DISKEVGASIKVQLKMEI 239

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 1903134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10649273/runat_02112005_091338_15562/app_query.fasta_1.1429
-DB=EST -QFMT=fastcap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human4.0.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10649273 @CGN 1 1 8076 @runat_02112005_091338_15562 -NCPU=6 -ICPU=3
-NO_MMAP -LARGQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOC
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_hic:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1118	90.2	870	5	BQ423651 AGENCOURT
2	1097	88.5	640	5	BQ636028 hnd03d11.y
3	1079	87.0	2284	3	AK045669 Mus muscu
4	1071	86.4	1622	3	AK011265 Mus muscu
5	930	75.0	852	5	BX391919 BX391919
6	926	74.7	1605	5	BCU30671 Mus muscu
7	908	73.2	701	2	BE740611 601595739
8	876	70.6	658	7	CF362328 829596 MA
9	838	67.6	637	7	CK941819 4065407 B

10	799	64.4	922	5	BQ961028	BQ961028 AGENCOURT
11	789	63.6	545	1	AV602901	AV602901
12	784	63.2	490	6	CB852881	CB852881 UI-CF-FNO
13	783	63.1	822	7	CF257246	CF257246 pha008_90
14	753	60.7	866	5	BUI27463	BUI27463 603114407
15	749	60.4	736	7	CK365185	CK365185 AGENCOURT
16	746	60.2	723	5	BUI261251	BUI261251 603502215
17	731.5	59.0	792	5	BUI403563	BUI403563 604138456
18	717	57.8	634	2	AM601179	AM601179 RCL-BT025
19	708	57.1	736	6	CA057753	CA057753 sea1rbp54
20	707.5	57.1	701	5	BUI621780	BUI621780 UI-H-FL1-
21	698	56.3	696	5	BUI302605	BUI302605 603739448
22	685	55.2	537	6	CB272391	CB272391 ma157912.
23	683	55.1	909	5	BX756548	BX756548 BX756548
24	682	55.0	548	7	CO880741	CO880741 Bovenen_09
25	671	54.1	682	2	BB043703	BB043703 BB043703
26	671	54.1	730	7	CN833245	CN833245 Oa.sp1bn
27	668	53.9	706	5	BUI202465	BUI202465 603949052
28	666	53.7	869	5	BX754527	BX754527 BX754527
29	647	52.2	919	7	CF407294	CF407294 CH3#047_E
30	646	52.1	484	1	AJ670918	AJ670918 AJ670918
31	646	52.1	878	5	BX776940	BX776940 BX776940
32	642	51.8	1173	6	CD508917	CD508917 CDA93-E05
33	638	51.5	1082	5	BX359023	BX359023 BX359023
34	625	50.4	506	2	BF415802	BF415802 UI-R-CA1-
35	614	49.5	861	5	BUI26158	BUI26158 603779906
36	611	49.3	1171	5	BUI261605	BUI261605 603501763
37	604	48.7	424	1	AA273921	AA273921 vb99c03_r
38	600	48.4	863	5	BUI376295	BUI376295 603808890
39	594	47.9	389	4	BM744822	BM744822 K-BST0018
40	591	47.7	812	5	BUI246489	BUI246489 603784202
41	589	47.5	357	5	BQ672554	BQ672554 AGENCOURT
42	581	46.9	704	5	BUI37284	BUI37284 603491570
43	576	46.5	826	7	CR444994	CR444994 CR444994
44	573	46.2	385	1	AJ647827	AJ647827 AJ647827
45	572.5	46.2	613	1	AA920105	AA920105 v225H05.r

ALIGNMENTS

RESULT 1
BQ423651 870 bp mRNA linear EST 23-MAY-2002
LOCUS BQ423651
DEFINITION AGENCOURT 7790948 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:6065828
5', mRNA sequence.

ACCESSION BQ423651 GI:21118966
VERSION BQ423651.1
KEYWORDS EST.

SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov

Tissue Procurement: ATCC/DC/DP
CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov

Plate: L1AM13342 row: 1 column: 21
High quality sequence stop: 710.

FEATURES

source
1..870
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6065828"


```

/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 72"
/notes="Organ: Vector: PCMV-SPOB6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2 kb. Library constructed by Life
Technologies."

```

ORIGIN

Alignment Scores:

```

Pred. No.: 3e-119 Length: 870
Score: 1118.00 Matches: 220
Percent Similarity: 98.22% Conservative: 1
Best Local Similarity: 97.78% Mismatches: 2
Query Match: 90.16% Indels: 2
DB: 5 Gaps: 0

```

US-10-649-273-2_COPY_176_414 (1-239) x BQ423651 (1-870)

```

Qy 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20
Db CTGTTGGCATTTGTTCAAGAGATTTCAGATTTCCTCTTGGAAGCTTTGGACATA 255
Qy 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGln 40
Db GCACACAGGTGACATCTTGACAAAGGTGCAGAGAAAGCTTTTAAATAAACATCCAGAG 315
Qy 41 CysSerThrMetSerGlyGlyLysAlaIleGlnHisLeuAlaGlnGlyAsnArgPhe 60
Db TGCTCCACCATGAGTGGTGGGAAGCCATGAACATTGGCCAAACAGAAATATATATT 375
Qy 316 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80
Db CATTTTGACATTAACCTCCCTTGATCTGCTAATAAATTGGATTTCCTTTTACTGCA 435
Qy 376 LeuGlnHisValThrAspLysIleIleMetLysGlnLysGlnLysGlnLysGlnLys 100
Db CTTCAACAGCTTACGATTAATAATGAAGAAAGAAAGAGAGATTTGAGAG 495
Qy 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 120
Db GGGCAAACTCTGCTTCACGACAGACATGCTGCCACAGTACAGCACCAATGGCATGT 555
Qy 496 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLeuGlnArgAspLeuPro 140
Db TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 615
Qy 121 CysThrAspAsnGlyIleMetIleAlaThrAsnGlyIleGlnArgLeuArgAlaGlyLeu 200
Db TGCACCTGAAATGCGCATTAAGATGCAATGGAATGAAGACCTACGTCGCTG 795
Qy 736 GlyIleLeuHisAspIleGlnGlyIleArgGlyLeu-ProLysCysPro-LeuGlyValAla 220
Db GGCATTTTTCATGACATTAAGAGCATCCGCTATGAACCAATAGTCCTCTTGAGAGTAG 855
Qy 796 spIleSerLys 223
Db 856 ACATATCAAAA 866

```

```

RESULT 2 BQ636028 640 bp mRNA linear EST 15-JUL-2002
LOCUS h03d11.y1 Human Retina cDNA (Un-normalized, unambigified): hd/he
DEFINITION Homo sapiens cDNA clone h03d11 5', mRNA sequence.
ACCESSION BQ636028

```

VERSION BQ636028.1 GI:21760487

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

AUTHORS

TITLE

Wistow, G., Bernstein, S.L., Wyatt, M.K., Ray, S., Behal, A.,

Touchman, J.W., Bouffard, G., Smith, D. and Peterson, K.

Expressed sequence tag analysis of human retina for the NEIRBank

Project: Retblind, an abundant, novel retinal cDNA and alternative

splicing of other retina-preferred gene transcripts

Mol. Vis. 8 (4), 196-204 (2002)

JOURNAL

MEDLINE

PUBMED

COMMENT

Contact: Wistow G

Section on Molecular Structure and Function

National Eye Institute

6/331, NIH, Bethesda, MD 20892-2740, USA

Tel: 301 402 3452

Fax: 301 496 0078

Email: graeme@helix.nih.gov

Plate: 03 row: d column: 11

Seq primer: M13RP1 reverse primer (ABI).

Location/Qualifiers

1. 640

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="hd03d11"

/tissue_type="Retina"

/dev_stage="Adult"

/lab_host="EMDHI08"

/clone_lib="Human Retina cDNA (Un-normalized,

unamplified): hd/he"

/note="Organ: Eye; Vector: pSPOB1; Neural retina tissue

was dissected from two 80 year old donors with no observed

eye disease. 100ug of total RNA was used for library

construction. A directionally cloned cDNA library in the

pSPOB1 vector (Life Technologies) was constructed at

Bioserve Biotechnology (Laurel MD) essentially following

the protocols of the Superscript Plasmid System full

details of which are contained in the manufacturer's

instruction manual (http://www.lifetech.com/). First

strand synthesis was carried out using a Not I

primer-adaptor

[5'-pGACTGATGTTCTGATGCGAGCGCGCCGCT(15-3')]. EST analysis

was performed on the unamplified library at the NIH

Intramural Sequencing Center (NISC)."

ORIGIN

Alignment Scores:

```

Pred. No.: 5.41e-117 Length: 640
Score: 1097.00 Matches: 209
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 88.47% Indels: 0
DB: 5 Gaps: 0

```

US-10-649-273-2_COPY_176_414 (1-239) x BQ636028 (1-640)

```

Qy 31 ArgArgLeuSerLeuIleLysHisProGlnCysSerThrMetSerGlyLysAlaIle 50
Db AGAAGACTTTCCTTAATATAACATCCAGAGTCTCCACATGAGTGTGGGAAGCCATA 62
Qy 51 GluHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHis 70
Db GAACATTGGCCCAACAAGAAATGAATTCATTGACATCAAAACCTCCCTGCATCAT 122
Qy 71 AlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMet 90
Db GCTAAATAATGTGATTTTCTTTTACTGAGCTTCAACACGTTACTGATAAATATATATG 182

```

QY	91	LYSLVSGILVSGILGILGILYLIEGLIULVSGILGILILEUSERSERLALALASPILE	110
Db	183	AAAAAGAAAAAGAGAGAGGTATTGAGAGGGGCAAATCTCTTCACAGACAGACATT	242
QY	111	AlaIarIhvalGIhIsthrMetAlaCysHIsleValILysaIghThriSarIaIaIle	130
Db	243	GCTGCACAGTACACACACACATGGCATGTCACTTGTGTAAGAAACACATCGGCTAATT	302
QY	131	LeuPheCysLVSGILaIrgAspIleuLeuProGlnaIhnaIaIValIleuValIAsaIrgILy	150
Db	303	CTGTTTGTGAACAAGACTTGTATCTCAAAATTAAGCAGTACTGGTGCATCTGGI	362
QY	151	GIYAlaIaIAsaIasnPheYrILleIrgaIgaIaIeUGILILEuThrAsnIaIarGln	170
Db	363	GGTGCGCAAGTAATTCTATATCCGACAGCTCTGGAATTTTAAACAACGCAACAG	422
QY	171	CysThrIleuLeuCysProProProIarGleuCYsthrAspAsnGILYIleMetIlealatrP	190
Db	423	TGCATTTTGTGTGTCCTCTCCACAGCATATGCACGTGAATAGGCAATTATGATTTGATGG	482
QY	191	AsnGILYIleGIaIrgIeuaIrgaIaGIlyLeuGILYILEuHIsaSPIleGIuGILYIleIarG	210
Db	483	AATGTATTTGAAGACTAGCTGCTGGCTTGGCATTTTACATGACATTAAGGCAATCCGC	542
QY	211	TyrGIuPProLYsCysProLeuGILYalAsPIIseIryLeGILValGILYIaIaIaIseIryIle	230
Db	543	TATGAACCAAAATGCTCTTGGAGTAGACATATCAAAAGAAGTTGGAGAAAGCTTCATA	602
QY	231	LYsValIProGlnIleuLYeMetGILuIle	239
Db	603	AAAGTACACATTTAAATAAGAGATA	629

RESULT 3				
AK045669		2284 bp	mRNA	linear HTC 03-APR-2001
LOCUS				
DEFINITION	Mus musculus adult male corpora quadrigemina cDNA, RIKEN full-length enriched library, clone:B230219017 product:similar to PUTATIVE SIALOGLYCOPROTEASE TYPE 2 [Homo sapiens], full insert sequence.			
ACCESSION	AK045669			
VERSION	AK045669.1 GI:26337528			
KEYWORDS	HTC; CAP trapper.			
SOURCE	Mus musculus (house mouse)			
ORGANISM	Mus musculus			
REFERENCE	1 Carninci,P. and Hayashizaki,Y. High-efficiency full-length cDNA cloning Mech. Enzymol. 303, 19-44 (1999) 99279253			
AUTHORS	2 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shiota,K., Itoh,M., Komo,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000) 20499374			
JOURNAL	3			
MEDLINE	11042159			
PUBMED				
REFERENCE				
AUTHORS	Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Komo,H., Akiyama,J., Nishi,K., Kitsuana,T., Teshito,H., Itoh,M., Suni,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kasihwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohnra,E., Wataniki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y. RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multichipillary sequencer Genome Res. 10 (11), 1757-1771 (2000) 20530913			
JOURNAL				
MEDLINE	11076661			
PUBMED				

REFERENCE	FEATURES
<p>4 THE RIKEN Genome Exploration Research Group Phase II Team and the PANTOM Consortium. Functional annotation of a full-length mouse cDNA collection Nature 409, 685-690 (2001)</p> <p>5 The PANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs Nature 420, 563-573 (2002)</p> <p>6 (bases 1 to 2284)</p> <p>Adachi, S., Aizawa, K., Akinura, T., Arakawa, T., Bono, H., Carrinoni, P., Fukuda, S., Furuno, M., Hanganai, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohnato, N., Okazaki, Y., Saico, R., Saichou, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.</p> <p>Direct Submission Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-501-9222, Fax: 81-45-503-9216)</p> <p>cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site for further details. URL: http://genome.gsc.riken.jp/ URL: http://fantom.gsc.riken.jp/. Location/Qualifiers</p>	<p>FEATURES</p> <p>SOURCE</p> <p>1..2284</p> <p>/organism="Mus musculus"</p> <p>/mol_type="mRNA"</p> <p>/strain="C57BL/6J"</p> <p>/db_xref="PANTOM.DB:B230219017"</p> <p>/db_xref="taxon:10090"</p> <p>/clone="B230219017"</p> <p>/sex="male"</p> <p>/tissue_type="corpora quadrigemina"</p> <p>/clone_id="RIKEN full-length enriched mouse cDNA library"</p> <p>/dev_stage="adult"</p> <p>221..1465</p> <p>/note="unnamed protein product; putative similar to PUTATIVE SIMILAR PROTEIN TYPE 2 [Homo sapiens] (SPTR O9H4B0, evidence: FASTY, 80%ID, 100%length, match=1242)"</p> <p>/codon_start=1</p> <p>/protein_id="BAC32450.1"</p> <p>/db_xref="GI:26337529"</p> <p>/translation="MLMLRTTAAIGIPKPKSKYVGLRPFVHPRLTSGKLVLIETSCDDTGAVVDENGVNIGALHSQTVLHTGVIIVPVAQIHRNIORIVETTLASCRIPSDLSAATIKGVIALSLGVLSFTQLVNPFKPEIPIHNHEAALITRLTNKVEPVPFLISGGHCLALVQGVSDPLFKSLDIAFGMLDVARRLSLIKHPESTMSGKALEOLAKDGNPHPTINPNNAKKSDSPGLOHITDKLTHKEKGEIEKGQIISADIDAAVQAHATCAKTRTHAILFCQKQKULSPANAVLVVSGVSNLITRKALRIVANAAQTLCPRLCTDNGIMIAWNGIIBRLRAGLVADIRIEPKPLGVDIRSEVAAATKIVRLKVAL"</p> <p>2262..2267</p> <p>/note="putative"</p> <p>2284</p> <p>/note="putative"</p>


```
source
1. 1622
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="PANTOM DB:2610001M19"
/db_xref="taxon:10090"
/clone="2610001M19"
/cisue_type="whole body"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="10 days embryo"
/notes="unnamed protein product: putative
similar to PUTATIVE STALOGLYCOPROTEINASE TYPE 2 [Homo
sapiens] (SPTK|Q9H4B0, evidence: FASTY, 80%ID, 100%length,
match=1242)"
/codon_start=1
/protein_id="BAB27506.1"
/db_xref="GI:12847276"
/translation="MLMLRTGAIPKPKKRVGFLRRFSVHPPTLSCHKVLGIET
SCDPTGAIVDEGTGVLGEALHSQTVHKTGIVPVAQQLHRENIQRIIVETLSAC
RITPSDLAIATTKIPGIALSFGVLSQVNOFKKPIIHMEALALIRLTK
VEPFLVLISGHCILALVQSGDFLLGKSLDAPGMLKVRRLSLIHGPGST
MSGKAIKBLAKDGRFHFTNPQNAKNCDSPTGIQHTDIXLTHKEKEGIEKE
QIISADIAAAVQHATKCHLAKTRHAIKFKQKULSPANAVLVSGVANSYIR
KALEIVANATCTLCPPRLCTDNGIMIAMNGIERLRGLGVHDVEDIRYKPCPL
GVDSIREVAEAIKVRILKML"
1605..1610
/notes="putative"
1622
/notes="putative"
```

polya_signal

polya_site

ORIGIN

Alignment Scores:

Pred. No.:	2.16e-113	Length:	1622
Score:	1071.00	Matches:	204
Percent Similarity:	91.21%	Conservative:	14
Best Local Similarity:	85.36%	Mismatches:	21
Query Match:	86.37%	Indels:	0
DB:	3	Gaps:	0

US-10-649-273-2_COPY_176_414 (1-239) x AK011265 (1-1622)

Qy 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20

Db 732 CTGTGGCATTAAGTCCAGAGTGTTCCTGATTCCTGCGGAGCTTTGGACANA 791

Qy 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40

Db 792 GCACACAGCGCAGATCTTGACACAGCTGACAGAACATCTTTATATCAACATCCAGAA 851

Qy 41 CysSerThrMetSerGlyLysValAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60

Db 852 TGTCTCAATGAGGTGAGAAAGCTATAGAACAGTGGCCAAAGACGGAATATGATTC 911

Qy 61 HisPheAspIleLysProPheLeuHisIleAlaLysAsnCyAspPheSerPheThrGly 80

Db 912 CATTACTACTCAATCCACTATGCAAGATGCTTAAGAAATGGCATTTTCTTCCAGGGA 971

Qy 81 LeuGlnHisValThrAspLysIleIleMetLysGlyLysGluGluGluGlyIleGluLys 100

Db 972 CTTCAACATATTAATGATTAAGCTATTAACACAGAAAGAAAGAAAGCAATGAGAG 1031

Qy 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 120

Db 1032 GAGCAAAATCTGTATCAGCTGACAGATTCGTCGCGTACAGATCCACAGCGTCC 1091

Qy 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140

Db 1092 CACCTTGGAAAGAACACATGCGCTATCTGTTTGGACAGAAATTTGCTCTCT 1151

Qy 141 GlnAsnAsnAlaValLeuValAlaSerGlyValAlaSerAspPheTyrlleArgArg 160

Db 1152 CCACCTAAGCAGATTAAGTTGTAATCTGAGGTGTTGCAAGTAACTGTGATCCGAAA 1211

Qy 161 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180

Db 1212 GCATTGGAAATGTCCCAATGCAACGAGTGACGTTGTGTGTCCACCTCCAGACTG 1271

Qy 181 CysThrAspAsnGlyIleMetIleAlaATTPAsnGlyIleGluArgLeuAlaGlyLeu 200

Db 1272 TGACCTGACAAATGCAATCATGATGATGCAATGGAATTAAGATTACGTGCGCTTG 1331

Qy 201 GlyIleLeuHisAspIleGluGlyIleArgTyrlleGluProLysCysProLeuGlyValAsp 220

Db 1332 GCGGTTTACATATGATGAGAGACATCCGATTAAGAACCAAAATGCTCTTGAAGTAC 1391

Qy 221 IleSerLysGluValGlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 239

Db 1392 ATATCCAGAGAACTGACAGAACTGCATTAAGTACCGGATTAATAATGACACTT 1448

RESULT 5

BX391919/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Contact: Genoscope

Genoscope - Centre National de Sequencage

2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE

Email: secref@genoscope.cns.fr, Web: www.genoscope.cns.fr

1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized.

division of Invitrogen. This sequence belongs to sequence cluster 1240.r

For more information about this cluster, see http://www.genoscope.cns.fr/cdna?c=CSDBA10352&E07_CS03317_1&c=1240.r

FEATURES

source

Location/Qualifiers

1. 852

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="CSDBX001YB02"

/cell_type="HELA CELLS COT 25-NORMALIZED"

/clone_lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"

/notes="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Alignment Scores:

Pred. No.:	2.28e-97	Length:	852
Score:	930.00	Matches:	180
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	75.00%	Indels:	0
DB:	5	Gaps:	0

US-10-649-273-2_COPY_176_414 (1-239) x BX391919 (1-852)

Qy 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20

Db	542	CTGTTGGCATAGTTCAAGGAGTTTCAATTTTCGCTTTGGAAATCTTTGGACATA	483
Qy	21	AlaProGlyAspMetLeuAspLysValAlaArgGluSerLeuIleYshIProGlu	40
Db	482	GCACGAGGTGACATGCTTGACACAGGTGGCAAGAGACTTTCTTAAATAACATCCAGAG	423
Qy	41	CysSerThrMetSerGlyGlyLysAlaIleGluIshLeuAlaYshGlnGlyAsnProPhe	60
Db	422	TGCTCCACCAAGAGTGTGGAAAGCCATGAGACATTTGGCCAAACAGAAATAGATTT	363
Qy	61	HisPheAspIleLysProProLeuIshIshAlaYshAsnCyAspPheSerPheThrGly	80
Db	362	CATTTTGACATCAACCTCCCTTGATCATGTATAAATTTTGATTTTCTTTTACTGGA	303
Qy	81	LeuGlnHisValThrAspLysIleIleMetLysLysGlyLysGlnGlnGlyIleGluLys	100
Db	302	CTTCACACGCTTACTGATMAATATATATGAAAAAGAAAAAGAGGAAGTATTGACAGG	243
Qy	101	GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleAlaThrValGlnHisThrMetAlaCys	120
Db	242	GGGCAATCTCTGCTTCACAGCAGACATTTGCTGCCACGTCACGACACACATGGCAGTT	183
Qy	121	HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro	140
Db	182	CATCTGTGGAAAAAGACACATCGGGCTATTCTGTTTGTAAAGCAGAGACTTGTTAACCT	123
Qy	141	GlnAsnAsnAlaValLeuValAlaSerGlyValAlaSerAsnPheTyrlleArgArg	160
Db	122	CAAAATATATGACGACTCGGTGCACTCTGTGGTGTGCGAAGATMACTTATATATCCGACGA	63
Qy	161	AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu	180
Db	62	GCTCTGGAAATTTTACAAACGCAACACAGTGCACCTTTGTGTGTCTCTCCACGACTA	3

REMARK	JOURNAL	TITLE
NIH-MGC Project URL: http://mgc.nci.nih.gov Contact: MGC help desk Email: cgabs-remail.nih.gov Tissue Procurement: Marcello Bento Soares, Ph.D. cDNA Library Preparation: M. Bento Soares, University of Iowa cDNA Library Arrayed by: The I.M.A.G.E. Consortium (IuNL) DNA Sequencing by: Institute for Systems Biology http://www.systemsbio.org contact: amadamsystemsbiology.org Anup Madan, Jessica Pahey, Erin Helton, Mark Ketterman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whitting	Direct Submission Submitted (20-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC) Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA	Clone distribution information can be found through the I.M.A.G.E. Consortium/IuNL at: http://image.lnl.gov Series: IRAX Plate: 66 Row: e Column: 10 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gl: 21312463 This clone has the following problem: frame shifted. Location/Qualifiers 1. 1609 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="IMAGE:1226118" /tissue_type="Thymus gland, mouse" /clone_id="Soares_thymus_2NBWT" /lab_host="DH10B" /note="Vector: pRT73-Pac"

RESULT 6	BC030671	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
REFERENCE	BC030671	1609 bp	mRNA	linear	HTC 19-NOV-2003			
REFERENCE	Mus musculus O-sialoglycoprotein endopeptidase-like 1, mRNA (cDNA clone IMAGE:1226118), containing frame-shift errors.							
REFERENCE	BC030671							
REFERENCE	BC030671.1	GI:21040459						
REFERENCE	HTC.							
REFERENCE	Mus musculus (house mouse)							
REFERENCE	Mus musculus							
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.							
REFERENCE	1 (bases 1 to 1609)							
REFERENCE	Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, P., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stopleton, M., Soares, M.B., Donald, M.F., Cavaant, T.L., Scheer, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Sacchini, P., Prange, C., Rana, S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullahy, S.J., Bosak, S.A., McMan, P.J., McGernan, K.J., Malek, J.A., Gumarate, P.H., Richard, S., Wotley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.N., Viallalon, D.K., Muzny, D.M., Sodergren, E.U., Lu, X., Gibbs, R.A., Fahey, J., Heltom, E., Kettelman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butlerfield, Y.S., Krzywinski, M.I., Skalka, U., Small, D.E., Schermer, A., Schein, J.E., Jones, S.J., and Matra, M.A.							
REFERENCE	human and mouse cDNA sequences							
REFERENCE	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)							
REFERENCE	22388257							
REFERENCE	12477932							
REFERENCE	2 (bases 1 to 1609)							
REFERENCE	Strausberg, R.							

Alignment Scores:			
Pred. No.:	1,65e-96	Length:	1609
Score:	926.00	Matches:	183
Percent Similarity:	81.59%	Conservative:	12
Best Local Similarity:	76.57%	Mismatch:	20
Query Match:	74.68%	Indels:	24
DB:	3	Gaps:	1
US-10-649-273-2_COPY_176_414 (1-239) x BC030671 (1-1609)			
Qy	1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIle	20	
Db	749 CTGTTGGCATTATAGTCCAAAGGTGTTCCGATTTCCTGCTCCCTTGGAAGTCTTTGGACATA	808	
Qy	21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu	40	
Db	809 GCACGAGCGACATGCTTGACAAAGGTGGCAAGAAAGACTTTCTTTAATCAACATCCAGAA	868	
Qy	41 CysSerThrMetSerGlyGlyLysAlaIleGluHisIleuAlaLysGlnGlyAsnArgPhe	60	
Db	869 TGTTCCTCAATGAGGTGGTGAAGAAAGTATATAGAACAGTTGGCCAAAGACGGAATAAGATT	928	
Qy	61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly	80	
Db	929 CATTTTACTATCAATCCACTATACAGATGCTAAGAAATTCGATTTTCTTTCACGGGA	988	
Qy	81 LeuGlnHisValThrAspLysIleIleMetLysGlyGlnLysGlnGlyIleGluLys	100	
Db	989 CTTCAACATATTACTGATTAAGCTAATTAACACACAGAGAAAAAGAAAGCATTGAGAAAG	1048	
Qy	101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCys	120	
Db	1049 GGGCAAAATCCCTGTCATCAGCTGCAGACACTTCCTCTGGCTGACAGCAACGACGCGC	1108	
Qy	121 HisIleuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuPro	140	
Db	1109 CACCTTGCGAAAAGAACATATGGCTATATCTGTTTTCAGACGAGAAAATTTGCTCTCT	1168	

Alignment Scores:	
Pred. No.:	3,07e-91
Score:	876.00
Percent Similarity:	93.68%
Best local Similarity:	86.84%
Query Match:	70.65%
DB:	7
length:	658
Matches:	16
Conservative:	13
Mismatches:	12
Indels:	0
Gaps:	0

US-10-649-273-2_COPY_176_414 (1-239) X CF362328 (1-658)

OY	1	LeuLeuAlaIeuValGlnGlyValSerAspPheLeuLeuLysSerLeuAspIle	20
Db	630	CTTTTGCGATTAGTGAAGAAGACTTCACGATTTCTCCTCTTGACAGCTTTGGACATA	571
OY	21	AlaProGlyAspMetLeuAspLysValAlaArgLeuSerLeuIleLysHisProGlu	40
Db	570	GCACCAGGTGACATCTGTGACAAAGGTAGCAAGAAGACTTTCTTAATAAACAATCCAGAG	511
OY	41	CysSerThrMetSerGlyGlyValSalAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe	60
Db	510	TGCTCCACCATGAGGTGGGAGGCCAATGAGCAATTTGTCCAAACAGGGAATAAGTTG	451
OY	61	HisPheAspIleLysProProLeuHisHisSalAlaLysAsnCysAspPheSerPheThrgly	80
Db	450	CATTTTGATTTCCAACCTCCCATGCAACGTGATATAAAATTTGATTTTCTTTTCTGCA	391
OY	81	LeuGlnHisValThrAspLysIleIleMetLysGluLysGluGlnGlyIleGluLys	100
Db	390	CTTCACACATGTTATGATNAGACATATTAATGCAAGAAAGGAAGAGATATTGAGAAG	311
OY	101	GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleAlaThrValGlnHisThrMetAlaCys	120
Db	330	GGGCAAACTCCTCTTCAGCTGCAGACATTCGTCTGCAGATCACAGACACAGTAGCCTGC	271
OY	121	HisIleuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro	140
Db	270	CATATTGCAAAAGAACCTATCGTGTATTCGTGTTTGCMAACAGAGACTTATTAATGT	211
OY	141	GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPhenylIleArgArg	160
Db	210	CAAGTATATGCAGTATGGTGTATCTGGAAGGTTCGCAAGTAACTTATATATACAAAA	151
OY	161	AlaIeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu	180
Db	150	GCTTTAGAGTGTGACAAATGCACACACATGCACTCTGTGTGCTCTCCCAACTA	91
OY	181	CysThrAspAsnGlyIleMetIleAlaTrp	190
Db	90	TGCACGTAAATAGGCATTATGATTCGATGG	61

RESULT 9	CK941819/c	CK941819	637 bp	mRNA	linear	EST 15-MAR-2004
LOCUS						
DEFINITION	4065407	BARC 10BOV	Bos taurus	CDNA clone 10BOV12_F24 3', mRNA		

ACCESSION	CK941819	
VERSION	CK941819.1	GI:45456199
KEYWORDS	EST.	
SOURCE	Bos taurus (cow)	
ORGANISM	Bos taurus	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi	

Animal and Natural Resources Institute
Bd19, 200 Km24, BARC-East, Beltsville, MD 20705, USA
Tel: 3015048416
Fax: 3015048414
Email: tadseanri.barc.usda.gov

Single pass sequencing. Bases called and trimmed with phred
0.000925 using options -trim_aln "-trim.fasta. Vector identified
by cross match using options -nmatch 12 -mnscore 18
Plate: 12 row: F column: 24
Seq primer: AGCGATTACCAATTTCACACAGS
High quality sequence stop: 637.

FEATURES	Location/Qualifiers
source	1. .637

```

source

1. .637
/organism="Bos taurus"
/mol_type="mRNA"
/strain="Holstein"
/db_xref="taxon:9913"
/clone="10BOV12_F24"
/sex="Male"
/tissue_type="Pooled"
/dev_stage="Multiple"
/lab_host="DH10B T1 phage resistant"
/clone_lib="BARC 10BOV"
/note="Organ: Small Intestine; Vector: pagen-1; Site: 1; ECoRV; Site 2: NotI; Equimolar amounts of mRNA extracted from proximal jejunums of 18 and 21 wk old steers, and distal ileums of 14 day old calves. proximal jejunum exposed to C. oncophora for 3 and 6 weeks, and distal ileum exposed to C. parvum for 7 days"

```

Alignment Scores:			
Pred. No.:	7,81e-87	Length:	6
Score:	838.00	Matches:	1
Percent Similarity:	92.47%	Conservative:	1
Best local Similarity:	83.26%	Mismatches:	0
Query Match:	67.58%	Indels:	0
DB:	7	Gaps:	0

```

OY 54 AlAlaYsGInGlyLysMaRgPheHISpHeaPpIleLysProPLeuHIShISaIaLysAaR 73
Db 632 GCCAAACAGGAAATAGATTGACATTTTGATTTCACACCTCCATCCAGAACGTGTAAAAAT 57
OY 74 CysApSeSerPheThrGlyLeuGlnHISvalThrAspLysIleIleMetLysLysGlu 93
Db 572 TGTGATTTTCTTTTCTTGACATTCACACGTTATGATGAAGATGATATGCAAAAGAA 51
OY 94 LysGIuGIuGIuIleGluLysGIuGlnIleLeuSerSerAlaAlaSpIleAlaIaThr 11
Db 512 AAAGACGAAGATTCGACGACGAGGCGAGTCTCTCTTCACGTCGACACATTCGTTGCG 45
OY 114 ValGlnHIShThrMetAlaCysHISLeuValLysArgThrHISArgAlaIleLeuPheCys 13
Db 452 GTGACGACACCGGTGGCTGCCACATTGCAGAAAGAACACATGTCTCTCTCTGTTTCC 39
OY 134 LysGlnArgAspLeuLeuProGlnIleAsnAsnAlaValIleuValAlaSerGIuGIuVala 15
Db 392 AAGCAGAGAGCGTTCTTACGTACGACAGATGAACGACGTACTGTTATCTCGAGGCGTCGA 33
OY 154 SerAsnPheTyrtIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThrIleu 17

```


Db 332 AGTAATTATATATCCGAAAGCCCTGGAATTTGTGACCAATGCAACACAGTGCACCTTTG 273

Qy 174 LeuCySPICProPrgArgLeuCySPthAspAenglylleMeTllealTrpAenglylle 193

Db 272 CTGTGCCCCCCCCCGAGCTTTGCACTGACCAACGGGTTATGATTCATGATGAGTGT 213

Qy 194 GUATGLeuAArgAlGlyLeuGlylleLeuHisAspIleGluGlylleArgTyrGluPro 213

Db 212 GAAAGACTAGCTGCTGGCTTGGGCAATTTTACACACAGAGGCAATCCGCTACGAAACA 153

Qy 214 LysCySPICProLeuGlyValAspIleSerLysGluValAlGlyValAspIleSerLysValPro 233

Db 152 AAATGCTCCCTTGGAGTATATATATCAAAAGAGTTGAGAGAGCTCTATTAAGATGCCA 93

Qy 234 GlnLeuLysMetGluIle 239

Db 92 AGATTAAAAATGAGATT 75

RESULT 10

BO961028 922 bp mRNA linear EST 21-AUG-2002

LOCUS AGENCOURT 8863711 NIH_MGC_18 Homo sapiens cDNA clone IMAGE:6423902

DEFINITION 5', mRNA sequence.

ACCESSION BO961028

VERSION BO961028.1 GI:22376506

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgsabp@remail.nih.gov
Tissue Procurement: DCTD/DP/GenZdar
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM2603 row: d column: 15
High quality sequence stop: 584.

FEATURES

source

1..922

location/Qualifiers

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:6423902"

/tissue_type="large cell carcinoma"

/lab_host="DH10B (phage-resistant)"

/clone_1lb="NIH_MGC_18"

/note="Organ: lung; Vector: pOT7; Site 1: XhoI; Site 2: EcoRI; XhoI sites using the following 5' adaptor: GGCAAGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC library."

ORIGIN

Alignment Scores:

Pred. No.: 4.6e-82 Length: 922

Score: 799.00 Matches: 179

Percent Similarity: 88.04% Conservative: 5

Best Local Similarity: 85.65% Mismatches: 16

Query Match: 54.44% Indels: 10

Gaps: 5 Gaps: 3

US-10-649-273-2_copy_176_414 (1-239) x BO961028 (1-922)

Qy 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20

Db 291 CTGTGGCATTAAGTTCAGAGAGTTTCAGATTTCTGCTTGGAAAGCTTTGGACATA 350

Qy 21 AlaProGlyLysPheLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40

Db 351 GCACAGAGTACATGCTTGCAGAGTGGCAAGAGACCTTTCTTTAAATTAACATCCAGAG 410

Qy 41 Cys-SerThMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60

Db 411 TGCNTCCACCATAGTGTGTGGAAAGCCATAGAACATTTGGCCAAACAGAAATAGATT 470

Qy 60 eHisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspSerPheThrGly 80

Db 471 TCATTTGACATCAAACTCCCTTGATCATGCTAAATAATTTGATTTCTTTTACTGG 530

Qy 80 yLeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGlyIleGlyLys 100

Db 531 ACTTCACACAGTTACTGATTAATTAATGAAGAAAGAAAGAGAGATTTAGAA 590

Qy 100 eGlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaC 120

Db 591 AGGCAAAATCTCTGCTTCAGACAGACATTTGCTGCCACAGTACAGCACAAAATGCGATG 650

Qy 120 sHisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPhe 140

Db 651 TCATCTTGTGAAAAAGAA-CATCGGCTATTTCTGTTTGTACAGAGACATTTGGTAC 709

Qy 140 OGlnAsnAsnAlaValLeuValAlaSer-GlyGlyValAlaSerAsnPheTyrIleArg 160

Db 710 TCAAAATTAATGACGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 769

Qy 160 rGlnAlaLeuGluIleLeuThr-AsnAlaThrGlnCysThr-LeuLeuCysProProArg 179

Db 770 AAACCTGGAATAATTTACAAACCCAGGAGCACTTTGTTGGCTCCCTCCCA 829

Qy 179 gLeuCySPthr-AspAenglylleMeTllealTrp-----AsnGlyIleGluArg---L 196

Db 830 ACTATGACATGATTAATGG---CATTTATGATGCTGCGGAGAGGAATTTGAAAAATA 886

Qy 196 euArgAlaGlyLeuGly 201

Db 887 CCNTGCTCGGCTTGGG 903

RESULT 11

AV602901/c 545 bp mRNA linear EST 27-NOV-2001

LOCUS AV602901 Bos taurus kidney fetus Bos taurus cDNA clone EIK1013A07

DEFINITION 3', mRNA sequence.

ACCESSION AV602901

VERSION AV602901.1 GI:9725227

KEYWORDS EST.

SOURCE Bos taurus (cow)

ORGANISM Bos taurus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.

AUTHORS 1 (bases 1 to 545)

TITLE Takasuga, A., Hirotsune, S., Itoh, R., Itoh, R., Itoh, R., Aso, H. and Sugimoto, Y.

ABSTRACT Establishment of a high throughput EST sequencing system using poly(A) tail-removed cDNA libraries and determination of 36,000 bovine ESTs

Nucleic Acids Res. 29 (22), E108 (2001)

21570554

11713328

CONTACT: Yoshikazu Sugimoto

Animal Genetics Division

Shirakawa Institute of Animal Genetics

Odakura, Nishigo, Niigata-shirakawa, Fukushima 961-8061, Japan

Tel: 81-248-25-5641

Fax: 81-248-25-5725

Email: kazusugi@cocca.ocn.ne.jp
Single pass sequencing.
This clone was obtained from a polyA-deleted cDNA library.

FEATURES

SOURCE

Location/Qualifiers

1..545
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/clone="EIK1013A07"
/tissue_type="kidney"
/dev_stage="fetus"
/lab_host="DH10B"
/clone_lib="Bos taurus kidney fetus"
/note="Vector: pZU1; Site 1: SalI; Site 2: NotI; Poly A
was deleted from a NotI site"

ORIGIN

Alignment Scores:

Pred. No.:	3,18e-81	Length:	545
Score:	789.00	Matches:	147
Percent Similarity:	90.11%	Conservative:	17
Best Local Similarity:	80.77%	Mismatches:	17
Query Match:	63.63%	Indels:	1
DB:	1	Gaps:	0

US-10-649-273-2_COPY_176_414 (1-239) x AV602901 (1-545)

OY 57 G1YAAATAGPHEHISPHASPILEYSPROFROLEUHHIS-ALALYSANCYASRPH 76
DB 544 GGAATATAGATTGATTTGATTTCCANCMTCCTGCAACGCGTCAAAATTTGATTT 485
OY 76 ESErPHERG1YLEUg1NHISVAlThrasp1ys1le1EMe1LySg1u1rSg1u1 96
DB 484 TTCTTTTGTGACCTTCAACGCTTATTTGATTAATGATTAATGCACAAAGAAAGAAAGCA 425
OY 96 UG1Y11EG1u1ySg1u1N1le1euSeSerAla1aAsp1leAla1aThrValG1NH1 116
DB 424 AGGTATCGAGCAGGCGGCAAGTCTCTTCAAGTGGGACATTTGCTGGCGGTCCAGCA 365
OY 116 sThMeAlaCySH1e1euVal1yArGThrH1sArGAla1le1euPheCy1ySg1u1r 136
DB 364 CACCGTGGCTCCCACTTGCACAAAGAAAGAAACACATCTGCTCTTCTTGGCAAGCAG 305
OY 136 GASp1eu1euP1roG1nAsnAsnAlaVal1euVal1AserG1yG1yVal1AserAsn 156
DB 304 AGGCTTCTTACATCGAGATGACGACATCTGCTGATCTGGAGCGTCCAGATTAATT 245
OY 156 eTy11eArGArGAla1eUg1u1leUThrasnAlaThrG1nCyS1Thr1eu1euCySP1 176
DB 244 ATATATCCGAAAGCCCTGGAATTTGTGACCAATGCAACACATGCACTTTGCTGCCCC 185
OY 176 oP1roP1oArG1eUyS1Thrasp1nG1y11eMe11eAla1TPAsnG1y11eG1u1rG1e 196
DB 184 GCCCCCAAGACTCTGCACTGACACACGCGCTNATGATTTGCAATGATGTTGAAAGACT 125
OY 196 UArGAlaG1y1eUg1y11e1euH1sAsp11eG1u1y11eArG1yG1u1P1ro1yS1CySP1 216
DB 124 AGGTGCTGGCTTGGGCAATTTTACACAAACAGAAAGCATTCGCTACGAAACAAAGTCC 65
OY 216 o1euG1yVal1Asp11eSer1ySg1u1yG1y1u1a1Ser11e1yS1Val1P1roG1n1eu1y 236
DB 64 TCTTGGAGTATATATCAAAAGATTGGAAGAGCTGCTATTAAGTCCCAAGATTAA 5
OY 236 sMet 237
DB 4 AATG 1
RESULT 12
LOCUS CB852881 490 bp mRNA linear EST 22-APR-2003
DEFINITION UI-CF-FNO-af0-b-03-0-UI_s1 UI-CF-FNO Homo sapiens cDNA clone
ACCESSION CB852881 UI-CF-FNO-af0-b-03-0-UI 3', mRNA sequence.

VERSION CB852881.1 GI:30047942

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE
AUTHORS Buzakytol; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 490)
Normalisation and subtraction: two approaches to facilitate gene
discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)

MEDLINE 97044477

PUBMED 8889548

COMMENT Contact: McCray, PB

McCrays Lab

University of Iowa

2024 University of Iowa Med Labs, Iowa City, IA 52242, USA

Tel: 319 356 4866

Fax: 319 356 7171

Email: paul-mccray@uiowa.edu

Tissue Procurement: Dr. M. J. Welsh, University of Iowa

cDNA library preparation: Dr. M. Bento Soares, University of Iowa

cDNA library Arrayed by: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Researchers may obtain clones from Research

Genetics (www.resgen.com) or from Open Biosystems

(www.openbiosystems.com).

Seq primer: M13 FORWARD

POLYA=No.

Location/Qualifiers

source

1..490

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-CF-FNO-af0-b-03-0-UI"
/tissue_type="Human Lung Epithelial cells"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-CF-FNO"
/note="Organ: Lung; Vector: pT773-Pac (Pharmacia) with a
modified polylinker; Site 1: Score 1; Site 2: Not 1;
UI-CF-FNO is a subtracted cDNA library derived from two
normalized human lung epithelial cell libraries (BN1 and
DU1) The library was subtracted according to
Bonaldi, Lennon and Soares, Genome Research, 6:791-806,
1996. Additional information, contact:
bento-soares@uiowa.edu
TAG_SEQ=None found"

ORIGIN

Alignment Scores:

Pred. No.:	1.05e-80	Length:	490
Score:	784.00	Matches:	148
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	63.23%	Indels:	0
DB:	6	Gaps:	0

US-10-649-273-2_COPY_176_414 (1-239) x CB852881 (1-490)

OY 69 H1SH1S1Ala1ySAnCYASrPHESErPHEThrG1y1eUg1NH1SVal1Thrasp1ys1le 88
DB 40 CATCATGCTTAAATTTGATTTTCTTTTACGGAATCTCAACGCTTACGATTAATA 99
OY 89 11eMe1LySg1u1ySg1u1yG1u1yG1y11eG1u1ySg1u1N1le1euSeSerAla1a 108
DB 100 ATATGAAAAAGGAAAAAGGAAAGTATTAAGAAAGGGAATTCCTTCAAGACGA 159
OY 109 ASp11eAla1aThrVal1G1NH1sThMeAlaCySH1eUVal1ySArGThrH1sArG 128
DB 160 GACATTTGCTCCACAGTACGACACACATGACATGATCTTTGGAAGAAACATCGG 219
OY 129 AlAl1eU1PheCy1ySg1u1rGArGAsp11eU1P1roG1nAsnAsnAlaVal1eUVala1a 148


```

/dev_stage="36"
/lab_host="DH10B"
/clone_lib="CSEQCH120"
/notes=Organ: limbs; Vector: pBluescript II KS(+); Site:1:
EcoRI; Site:2: NotI; Modification of pBluescript II KS(+):
[Stratagene] vector to accommodate cDNA produced with the
T-primed protocol (Construction of uni-directionally
cloned cDNA libraries from messenger RNA for improved 3'
end DNA sequencing by Glenn Fu, et al. U.S. Patent #
6,387,624). Cut pBluescript II KS(+) with NotI and EcoRI.
Ligate in double stranded adaptor containing BspI and
BamHI sites [5'ggccggcgagcccgagtcgcaaaaaag]
[5'aattcttttcggatccgggctcgacgc]

```

ORIGIN

Alignment Scores:

Pred. No.:	9.58e-77	Length:	866
Score:	753.00	Matches:	149
Percent Similarity:	78.81%	Conservative:	37
Best Local Similarity:	63.14%	Mismatches:	50
Query Match:	60.73%	Indels:	2
		Gaps:	0

US-10-649-273-2_COPY_176_414 (1-239) x BU127463 (1-866)

```

1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlySerLeuAspIle 20
25 ATCTTGCGACGATGACGACGAGCTTCAGATTTCTTCTGCGACAGTCCATAGTATA 84
21 AlaProGlyAspMetLeuAspIleValAlaArgArgLeuSerLeuIleLysHisProGlu 40
85 GCACAGAGTGACATCTGATGATAGGTAGACAGAGCTCTTTAGTAGAAGCAGCCGGAG 144
41 CysSerThrMetSerGlyGlyValAlaIleGlnHisLeuAlaLysGlnIleAsnAlaGpHe 60
145 TGCCACGCGCATGGCTGGGGGAGGCAATAGAGCACTGCTCAAAACCGAGAGCTGGCA 204
61 HisPheAspIleLeuProPheLeuHisAlaLysAsnCysAspPheSerPheThrGly 80
205 CAGTACACTTCAGCTTCCTCCATGCAACAGTATCGTACGTGATTTCTTCTCGGA 264
81 LeuGlnHisValThrAspLysIleIleMetLysLysGlnLysGlnLysIleGluLys 100
265 CTTGAGACCTTGTGCAACAAGCCATTCTTCAGAAAGAAAGAAAGATATTCAAGAA 324
101 GlyIleIleLeuSerSerAlaAlaAspIleAlaIleAlaThrValGlnHisThrMetAlaCys 120
325 GGGGAAATCTGTCCTGCGGTTAAGACATCGCTGCTGTCACAGCATGTAGTGGCTGCT 384
121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuPro 140
385 CATATTATTCAGCGGACACACCGAGCCATGCTTCTTCGATGAGTAAAAACGACATATTATTA 444
141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheThrIleArgArg 160
445 CCAAAAACCTGCAACTCTGTTCTATCAGAGAGAGTTGCAAGTATACGATATACAGAAAA 504
161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProAlaGlu 180
505 GGACGTGACAGACTCTGGC-AATGCAAAACGGTTTGTCTTTCTGCTCTCTCCACAGGCTG 563
181 CysThrAspAsnGlyIleMetIleAlaIleArgAsnGlyIleGlnArgLeuArgAlaGlyLeu 200
564 TGCACCGATTAATAGGTATGATTCATGAAATGGATTAAGAAAGTTGGCGTGAAGATGT 623
201 GlyIleLeuHisAspIleGlnGlyIleArgGlyGlnProLysCysProLeuGlyValAsp 220
624 GGTATATA-TACAGTACGTATGCGCATCCCTATGAAACAAAGCTCCCTTGAATATGAT 682
221 IleSerLysGlnValGlyGlnAlaSerIleLysValProGlnLeuLys 236
683 ATTTCAAAAGAGTTGAAGAGATTCCATCAGAGTCCCAAGACTAAGG 730

```

RESULT 15

CK365185 736 bp mRNA linear EST 23-DEC-2003
 LOCUS AGENCOURT 17157291 NIH_MGC 231 Rattus norvegicus cDNA clone
 DEFINITION IMAGE:7097620 5', mRNA sequence.
 ACCESSION CK365185
 VERSION CK365185.1 GI:40331120
 KEYWORDS EST.
 SOURCE Rattus norvegicus (Norway rat)
 ORGANISM Rattus norvegicus
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE

1 (bases 1 to 736)
 NIH-MGC <http://mgs.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Daniela S. Gerhard, Ph.D.
 Office of Cancer Genomics
 National Cancer Institute / NIH
 Bldg. 31 Rm10A07 Bethesda, MD 20892
 Email: cgabs-remail.nih.gov

Tissue Procurement: Howard Jacobs
 cDNA Library Preparation: Express Genomics
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
 Plate: LLM14944 row: d column: 02
 High quality sequence stop: 736.

FEATURES

source

1..736
 location/Qualifiers
 /organism="Rattus norvegicus"
 /mol_type="mRNA"
 /db_xref="taxon:10116"
 /clone="IMAGE:7097620"
 /tissue_type="lung, pooled"
 /lab_host="DH10B Tona"
 /clone_lib="NIH_MGC_231"
 /note="Organ: lung; Vector: pExpress-1; Site:1: EcoRV;
 Site:2: NotI; RNA obtained from pooled lung tissue from a
 mix of male and female animals at 8 wk old. Tissues were
 snap-frozen and kept at -80C for two days before RNA
 extraction and purification (Tri-reagent method). cDNA was
 primed using oligo-dT primer:
 5'-pGACTAGTCTTAAATCGCGAGCGCCGCCCT(T)25-3' and cloned into
 the EcoRV/NotI sites of pExpress-1. Size-selection >1.4kb
 resulted in an average insert size of 2.3 kb. This primary
 library is not normalized (normalized primary library is
 NIH_MGC 232) and was constructed by Express Genomics
 (Frederick, MD). Note: this is a NIH_MGC library."

ORIGIN

Alignment Scores:

Pred. No.:	2.22e-76	Length:	736
Score:	749.00	Matches:	141
Percent Similarity:	91.02%	Conservative:	11
Best Local Similarity:	84.43%	Mismatches:	15
Query Match:	60.40%	Indels:	0
		Gaps:	0

US-10-649-273-2_COPY_176_414 (1-239) x CK365185 (1-736)

```

73 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 92
1 AACTGATATTTCTTTTACGGGACCTTCAACATGTACCCGATAGCTAATATACACACAG 60
93 GlnLysGlnGlnGlyIleGlnLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaAla 112
61 GAAAAAAGAGAGCATTTGAGAGGGGCAATCTCTTCATCAGCGGACGACATTTGCTGCT 120
113 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 132

```

```

Db      121 GCGTACGACGCAACAGCGTCCACCTTGCGAAAAGAACACATCGTCTATTCTGTT 180
Qy      133 CysIysGlnArgAspLeuLeuProGlnAsnAlaValLeuValAlaSerGlyGlyVal 152
Db      181 TGCCAGCAGAAAATTGCTATCTCCAGCTAACGCACTATTAGTTGTCTGAGAGTGT 240
Qy      153 AlaSerAsnPheTyrIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThr 172
Db      241 GCAAGTAACCTGTACATCCGAGAGCATTGGAATTGTAGCAAAATGCAACACATGCACT 300
Qy      173 LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaTyrAsnGly 192
Db      301 TTGTGTGTCTCCCTCCGAGCTGTGCACCTGACAAATGATCATGATGATGCAATGGA 360
Qy      193 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrGlu 212
Db      361 ATTGAAAGATTACGTCGTGCTTGGGCAATTTACATGATGTAGAGACATCCGATACGA 420
Qy      213 ProIysCysProLeuGlyValAspIleSerIysGluValGlyGluAlaSerIleIysVal 232
Db      421 CCAAAATGTCTCTCGAATAGACATATCCAGAGAAAGTTGCAAGAGCTGCCATAAAAGTA 480
Qy      233 ProGlnLeuIysMetGluIle 239
Db      481 CCACGATTAAAATGACACTT 501

```

Search completed: November 11, 2005, 02:14:17
 Job time : 3392.57 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: November 10, 2005, 16:54:39 ; Search time 169.638 Seconds
(without alignments)
2305.321 Million cell updates/sec

Title: US-10-649-273-2_COPY_176_414
Perfect score: 1240
Sequence: 1 L1ALVGVSDFLILGKSLDI.....DISKVGESAIKVPOLKMEI 239

Scoring table: BLOSUMP62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=x1p
-Q=/cgn2_1/USFTO.spool.p/US10649273/runat.02112005.091339.15581/app.query.fasta_1.1429
-DB=Issued_Patents_NA -OFMT=faetap -SUFFIX=p2n.rn1 -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blissum62 -TRANS=human40.cdd1
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptlo -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10649273 @CNC 1.1 177 @runat.02112005.091339.15581 -NCPU=6 -ICPU=3
-NO_MAP -LARGEOUTERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARM TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS.COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfile1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1240	100.0	1416	4	US-09-774-528-177
2	1240	100.0	1526	4	US-10-067-443-23
3	1240	100.0	2197	4	US-10-067-443-1
4	1213.5	97.9	1387	4	US-10-067-443-21
5	725	58.5	14364	4	US-10-067-443-20
6	248.5	20.0	94750	4	US-09-596-002-38
7	247	19.9	1053	4	US-09-540-236-806
8	231	18.6	1059	4	US-09-252-991A-884
9	231	18.6	1206	4	US-09-252-991A-801
10	224	18.1	1830121	4	US-09-557-884-1
11	224	18.1	1830121	4	US-09-643-990A-1
12	221.5	17.9	996	4	US-09-902-540-6612

C	13	221.5	17.9	2582	4	US-09-902-540-503	Sequence 503, App
	14	212	17.1	1074	4	US-09-543-681A-2341	Sequence 2341, App
	15	205	16.5	1032	4	US-09-489-039A-2050	Sequence 2050, App
	16	200	16.1	1315	1	US-08-087-797-1	Sequence 1, Appli
	17	197	15.9	1092	4	US-09-107-532A-2955	Sequence 2955, App
	18	190	15.3	1008	3	US-08-987-121A-5	Sequence 5, Appli
	19	190	15.3	1011	3	US-08-987-121A-3	Sequence 3, Appli
	20	189	15.2	822	4	US-09-710-279-727	Sequence 727, App
C	21	189	15.2	3993	4	US-09-710-279-3985	Sequence 3985, App
	22	187	15.1	1155	4	US-09-602-777A-147	Sequence 147, App
C	23	186.5	15.0	4403765	3	US-09-103-840A-2	Sequence 2, Appli
C	24	186.5	15.0	4411529	3	US-09-103-840A-1	Sequence 1, Appli
	25	186	15.0	1006	3	US-08-961-083-51	Sequence 51, Appli
	26	186	15.0	1006	4	US-09-536-784-51	Sequence 51, Appli
	27	186	15.0	1007	3	US-09-134-001C-1072	Sequence 1072, App
C	28	186	15.0	10974	3	US-08-961-527-214	Sequence 214, App
	29	184	14.8	1011	3	US-09-066-512-1	Sequence 1, Appli
	30	182	14.7	1011	4	US-09-583-110-2196	Sequence 2196, App
	31	182	14.7	1663	4	US-09-620-312D-6	Sequence 6, Appli
	32	181.5	14.6	432	4	US-09-328-352-261	Sequence 261, App
	33	180.5	14.6	1664976	4	US-08-916-421B-1	Sequence 1, Appli
	34	180.5	14.6	1664976	4	US-09-692-570-1	Sequence 1, Appli
	35	178	14.4	640681	4	US-09-790-988-1	Sequence 1, Appli
	36	177	14.3	1011	4	US-09-107-433-1618	Sequence 1618, App
	37	174.5	14.1	1230025	4	US-09-198-452A-1	Sequence 1, Appli
	38	174.5	14.1	1230230	4	US-09-438-185A-1	Sequence 1, Appli
	39	174	14.0	1026	3	US-09-149-624-1	Sequence 1, Appli
C	40	173	14.0	1101	4	US-09-134-000C-1551	Sequence 1551, App
C	41	170	13.7	15249	4	US-08-956-171B-102	Sequence 102, App
C	42	170	13.7	15249	4	US-08-781-986A-102	Sequence 102, App
C	43	168	13.5	3064	3	US-09-221-017B-794	Sequence 794, App
C	44	159.5	12.9	42325	4	US-08-311-731A-131	Sequence 131, App
C	45	145.5	11.7	3215	4	US-09-710-279-3566	Sequence 3566, App

ALIGNMENTS

RESULT 1
US-09-774-528-177
Sequence 177, Application US/09774528
Patent No. 6743619
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Zhou, Ying
APPLICANT: Goodrich, Ryle
APPLICANT: Liu, Chenghua
APPLICANT: Asundi, Vinod
APPLICANT: Ren, Feiyang
APPLICANT: Zhang, Jie
APPLICANT: Zhao, Qing A.
APPLICANT: Yang, Yonghong
APPLICANT: Xue, Aidong J.
APPLICANT: Wehrman, Tom
APPLICANT: Wang, Jian-Rui
APPLICANT: Wang, Dunrui
APPLICANT: Drmanac, Radoje T.
TITLE OF INVENTION: No. 6743619el Nucleic Acids and
TITLE OF INVENTION: Polypeptides
FILE REFERENCE: 802
CURRENT APPLICATION NUMBER: US/09/774,528
CURRENT FILING DATE: 2001-01-30
NUMBER OF SEQ ID NOS: 441
SOFTWARE: pt_fl_genes Version 2.0
SEQ ID NO 177
LENGTH: 1416
TYPE: DNA
ORGANISM: Homo sapiens
FEATURES:
NAME/KEY: CDS
LOCATION: (205)..(1305)
US-09-774-528-177
Alignment Scores:

	:	FILE REFERENCE:	D0073 NP	
	:	CURRENT APPLICATION NUMBER:	US/10/067,443	
	:	CURRENT FILING DATE:	2002-02-05	
	:	PRIOR APPLICATION NUMBER:	US 60/266,518	
	:	PRIOR FILING DATE:	2001-02-05	
	:	PRIOR APPLICATION NUMBER:	US 60/282,814	
	:	PRIOR FILING DATE:	2001-04-10	
	:	NUMBER OF SEQ ID NOS:	71	
	:	SOFTWARE:	Patentin version 3.0	
	:	SEQ ID NO 1		
	:	LENGTH:	2197	
	:	TYPE:	DNA	
	:	ORGANISM:	Homo sapiens	
	:	FEATURES:		
	:	NAME/KEY:	CDS	
	:	LOCATION:	(231)..(1472)	
	:	US-10-067-443-1		
	:	Alignment Scores:		
	:	Pred. No.:	2,07e-157	
	:	Score:	1240.00	
	:	Percent Similarity:	100.00%	
	:	Best Local Similarity:	100.00%	
	:	Query Match:	100.00%	
	:	DB:	4 Gaps: 0	
US-10-649-273-2_COPY_176_414 (1-239) x US-10-067-443-1 (1-2197)				
QY	1	LeuleuAlaleuValGInGIyValseRAsPheLeuleuGIyLSerSeIeunspile	20	
Db	756	CTGTGGCATTAGTTCAAGAGATTCACAGTTTTCTCGCTTGGAAAGCTTTGGACATA	815	
QY	21	AlAProGIyAappMetleuAsplysValAlaaGAggleuSerleuIlleyshIsProGIU	40	
Db	816	GCACAGGTGCATGCTTGACAAGGTGGCAGAAGACTTTCTTATATAAACATCCAGAG	875	
QY	41	CysSerThrmSetserGIyGIyLysAlallegluhIsleuAlalyseGInGIyAnArgPhe	60	
Db	876	TGCTTCACCATGAAGTGTGGAGAACCATAGAACATTTGGCCAACAGAAATAGATT	935	
QY	61	HisPheaSpIIelEysProProleuhIsHIsAlalySaSCyASpPheSerPheThnGIy	80	
Db	936	CATTTTGACATCAAACTCCCTTGATCATGCTTAATAATGTGATTTTCTTTACTGA	995	
QY	81	LeuGlNhIsvalIThrAsPlysllellewecLySLySGluYegluGluGlyllegluYS	100	
Db	996	CTTCAACACGTTACTGATTAATAATATATGAAAAAGGAAAAGAGAGATTCAGAG	1051	
QY	101	GlyGlnIleleuseSerAlaalaspIleAlaalathValGIhIsPhmetlaIcYS	120	
Db	1056	GGGCAAAATCCGTCTTCAGCACAGACAACTGTCGCACAGTACAGACAAATGCATGT	1111	
QY	121	HisleuVallysaRGThrHIsaRGAlalIleleuPheCySLySGlnARgAspleuEuPro	140	
Db	1116	CATCTTGGAAGAAAGACACATCGGCTATTCTGTTTGTAAACAGAGACTGTGTTACT	1177	
QY	141	GlnAsnAsnalavalLeuValAlasergIyValAlaseRasnPhetryllleaRG	160	
Db	1176	CAAATATATGACGATACGTGTCATCTGTGTGTGTGCAAGTAACTTCTATATCCGACA	1231	
QY	161	AlAleuGIuIlleuThrAsnalathnGIncySThrleuEuCyProProProArgleu	180	
Db	1236	GCTCTGGAAATTTTAAACAAGCAACACAGTGCACCTTGTTGTCTCTCCACAGCTA	1291	
QY	181	CysThraSPaengIylEmetIlealTRPaengIyllegluARgLeuARGlaGIyeu	200	
Db	1296	TGCATCATATATGGCATTATATGATTCATGGAATGTGATTTGAAAGACTAGCTGTGCTTG	1351	
QY	201	GIyIleleuHIsaspIlegluGIyIleatrgTYrguProLYScysBroleuGIyAlaap	220	
Db	1356	GGCATTATTAACATGACATGAAGGACCTCGCTATGAAACCAAAATGTCTCTTGGAGTAGAC	1411	
QY	221	IleserLySGluValGIyGluAlaserIllelySValProGInleuLYMetGIuIle	239	

Db 1416 NATCAAGAGAGTTCGAGAGCTTCATTAAGTACACATTAAATGAGATA 1472

RESULT 4
US-10-067-443-21
; Sequence 21, Application US/10067443
; Patent No. 6642041
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN
; TITLE OF INVENTION: SPINAL CORD, MP-1
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO: 21
; LENGTH: 1387
; TYPE: DNA
; ORGANISM: homo sapiens
US-10-067-443-21

Alignment Scores:
Pred. No.: 3.87e-154 Length: 1387
Score: 1213.50 Matches: 238
Percent Similarity: 90.15% Conservative: 0
Best Local Similarity: 90.15% Mismatches: 1
Query Match: 97.86% Indels: 25
DB: Gaps: 1

US-10-649-273-2_COPY_176_414 (1-239) x US-10-067-443-21 (1-1387)

QY 1 LeuLuuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20
Db 549 CTTGTTGGCATTAGTTCAAGAGAGTTTCAGATTTCCTGTTCTTGAAAGCTTTTGACATA 608
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgSerLeuIleLysHisProGlu 40
Db 609 GCACGAGGAGATCGTTTGACAAGGTGGCAGAAGACTTTCTTAATTAATCAATCCAGAG 668
QY 41 CysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnLysAsnArgPhe 60
Db 669 TGCTCCACCATGATGGTGGGAAAGCCATPAGACATTGGCCAACAGAAATAGATT 728
QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnGlyAspPheSerPheThrGly 80
Db 729 CATTTTGACATTAACCTCCCTTGCAATGCTTAATAATTTGATTTTCTTTTACTGGA 788
QY 81 LeuGlnHisValThrAspLysIleIleMetLysGlyLysGlnGlyLysIle----- 98
Db 789 CTTCAACACGTTACTGATTAATAATTAATGAAGAAAAGGAGAAAGCATATATTCTA 848
QY 98 ----- 98
Db 849 ATTGATAAGTTGAACAGATAATATTCTCGATTGGCTTAATAATAGCTCATTTTC 908
QY 99 -----GluLysGlyGlnIleLeuSerSerLysAlaAspIleAlaIleThrValGln 115
Db 909 TGCAGGTATGAGAAAGGGGCAAAATCCCTGCTTCGACGACAAACATTGCGCACGTAACG 968
QY 116 HisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGln 135
Db 969 CACACAAATGGCATGTCATCTTTGTGAAAAGAACATCGGGCTATTCTGTTTGTAAACAG 1028
QY 136 ArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsn 155
Db 1029 AGAGACTGTACTCTCAAAATATATGACGATAGCTGTTGCATCTGTGTGGTGGCAAGTAAC 1088
QY 156 PheTyrIleLeuArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCys 175

```
Db 1089 TTCTAATATCCGACAGCTCTGAAATTTTAAACAACGACACAGTGCATTTGTTGTGT 1148
Qy 176 ProProProArgLeuCystrhAspAsnGlyLeuMetIleAtpAsnGlyLeuGluArg 195
Db 1149 CTTCTCCACGACATACGATGATTAAGCATTAATGATGATGATGAAGTATGTAAGA 1208
Qy 196 LeuArgAlaGlyLeuGlyIleLeuHiseAspIleGluGlyIleArgYrGluProlysCys 215
Db 1209 CTACGCTGCTGGCTTGGCATTTTACATGACATGACATGACATCCGCTATGAACCAAAAT 1268
Qy 216 ProLeuGlyValAspIleSerLysGluValGlyIleAsnIleLysValProGluLeu 235
Db 1269 CCTCTTGAGATGACATATATCAAAAGAGTTGAGAGAGCTTCATTAAGATCACCAATTA 1328
Qy 236 LysMetGluIle 239
Db 1329 AAAATGAGATA 1340

RESULT 5
US-10-067-443-20
; Sequence 20, Application US/10067443
; Patent No. 6642041
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE, ENCODING A NOVEL, METALOPROTEASE HIGHLY EXPRESSED I
; TITLE OF INVENTION: SPINAL CORD, MP-1
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 20
; LENGTH: 14364
; TYPE: DNA
; ORGANISM: homo sapiens
US-10-067-443-20

Alignment Scores:
Pred. No.: 3 21e-86 Length: 14364
Score: 725.00 Matches: 186
Percent Similarity: 32.86% Conservative: 0
Best Local Similarity: 32.86% Mismatches: 0
Query Match: 58.47% Indels: 380
Gaps: 3
DB: 4

US-10-649-273-2_COPY_176_414 (1-239) x US-10-067-443-20 (1-14364)
Qy 29 ValAlaArgArgLeuSerLeuIleLysHisProGluCysSerThiMeSerGlyLys 48
Db 11840 GTGGCAAGAGACCTTCTTTATATAAATCATCCAGAGTCTCCACCATGATGATGGAAA 11899
Qy 49 AlaIleGluHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProLeu 68
Db 11900 GCCATAGAACATTTGGCCAAACAAAGAAATGATTTCAATTTGACATCAAACTCCCTG 11959
Qy 69 HisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThiPheAspLysIle 88
Db 11960 CATCATCTAAATAATTTGATTTTCTTTTACTGACTTCAACAGCTTACTGATATAATA 12019
Qy 89 IleMetLysLysGluLysGluGlu-----GlyIleGluLysGlyGlnI 96
Db 12020 ATATATGAAAAAGAAAAAGAGAAAGTATATTCTTAATTAGTAAGTTGAACAGATAAT 12079
Qy 97 -----GlyIleGluLysGlyGlnI 103
Db 12080 ATTCCTGATTTGCTCTTAATAATAGCTGCTCATTTCTGACAGGTATTTGAAGGGCCAAAT 12139
Qy 103 eLysSerLysAlaAspIleAlaIleThrValGlnHisThrMetAlaCysHisLeuVal 123
```

```
Db 12140 CTTGCTTTACAGACAGACATTTGCTGCCACAGTACAGCACAAATGGCATGCTATCTTGT 12199
Qy 123 LysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuProGlnAsnAs 143
Db 12200 GAAAAAACAATCGGGCTATTTCTGTTTGTAGAGAGAGACTGTCTTCAAAATTA 12259
Qy 143 AlaValIleu----- 146
Db 12260 TGCAGTACTGTAAGTTTATCTCATTTTATAGTAATAGTTACCTTGCATATGTTC 12319
Qy 146 ----- 146
Db 12320 TTTTTCACAAAGACCTTGACCTTGTTGTTAGATGAACAGATCTTATGCTTATGCTAG 12379
Qy 146 ----- 146
Db 12380 CCTGACAGTATGAATAATTATGACAGATAGAAAAGCTAACAGCCATTTCTTGTACTAGTT 12439
Qy 146 ----- 146
Db 12440 TGCTAGCTTTATGAGACAGCTGTATAGCTTCTATGSCACATTAAGTAAATTTTGCATCTT 12499
Qy 146 ----- 146
Db 12500 CTTGTTGATTTAAAAAGAGGCTTACAAATTAAGAAAGTAATGACAGTAACCTGCTATCACT 12559
Qy 146 ----- 146
Db 12560 ATTTTGAAGAAATAGTGTGATTTCTTCATCTTGTATGAATCCCTTGTGTTGTTGT 12619
Qy 146 ----- 146
Db 12620 TTTTAAATAGACCACTCAATTTAGACAGTGGAGGTGTATTCACACTTCCTGACACT 12679
Qy 146 ----- 146
Db 12680 AATGTTGATTAAGTTCTGATTAATCCATATTTGTACAGACCAAAATCCCTTAATGTG 12739
Qy 146 ----- 146
Db 12740 CTTAAAGCCTTGACAAACATCCTGTTTAACTGTATCTTAACTTTATCAATTTAAAT 12799
Qy 146 ----- 146
Db 12800 TATTAACTAAGTGGAAAAATGTTTAATGTAAGTAATTCATATGATGAATTTTACATGG 12859
Qy 146 ----- 146
Db 12860 ATATCAAGAAATATTTTTCAGAGTTATGTAGTAATAATGCACAAATAATAAAATTTTC 12919
Qy 146 ----- 146
Db 12920 AGGGTCTAAATATAGTACTAGATTTGAATTAATTAATAATTAATTTAGATGAAGGT 12979
Qy 146 ----- 146
Db 12980 TGAAGAAATATACAAAAATGCTAGTAATGTTTGTATGCTATTAAGATTTATTAATTA 13039
Qy 146 ----- 146
Db 13040 TTTTCTTTCCAAATTTTATTAATAGATATGTCATGTCGCATTAACCATCTCAAAAT 13099
Qy 146 ----- 146
Db 13100 GGGATGTTTATTAATGTTTAATGCTGATATTTTCTCCAGTTTAATTAGACCTTGT 13159
Qy 147 -----ValAlaSerGly 151
Db 13160 TCATATCCATATATGATAGTATTTTGTGTTTCTCAATTCCTTCAAGTTGCACTGATGG 13219
Qy 151 YValAlaSerAsnPheYrIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnC 171
```



```
Db 13220 TGTGGCAGTAACCTTCTATATCCGACAGCTCTGGAATTTTAACAAACGACACAGTCG 13279
QY 171 sThleuLeuCySPProProProlaLeuCySPThraspaSnglyIleMeIleIaIaTrp-- 190
Db 13280 CACTTGTGTGTCTCTCTCCAGACTATGACATATATGACATATGATTTGATTTGATGGA 13339
QY 190 ----- 190
Db 13340 AGCCACAGATATACGTCTTCACTCATATATATATTAATTAATGTCATTATATCAT 13399
QY 190 ----- 190
Db 13400 ACTAAGCTTCTCTCTTCCAGACTTGTGAGCTATGATTTATTTAANGCTTCTATATTA 13459
QY 191 --AenglyIleGlyIleArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlyIleA 210
Db 13460 GGAATGGATTAAGAAAGATACGTGCTGGCTTGGCAATTTTACATGACATAGAGGATCC 13519
QY 210 rGlyrGlyProLys 214
Db 13520 GCTATGAACCAAG 13533

RESULT 6
US-09-596-002-38
; Sequence 38, Application US/09596002
; Patent No. 6632636
; GENERAL INFORMATION:
; APPLICANT: Lagace, Robert, E.
; APPLICANT: Patterson, Chandra
; APPLICANT: Berg, Kim, L.
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME
; FILE REFERENCE: PM-0008-4 US
; CURRENT APPLICATION NUMBER: US/09/596,002
; PRIOR FILING DATE: 2000-06-16
; PRIOR APPLICATION NUMBER: 60/140,121
; PRIOR FILING DATE: 1999-06-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PERL Program
; SEQ ID NO 38
; LENGTH: 94750
; TYPE: DNA
; ORGANISM: M. catarrhalis
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: incyte template ID No. 6632636 38
; PUBLICATION INFORMATION:
; US-09-596-002-38

Alignment Scores:
Pred. No.: 3,02e-20 Length: 94750
Score: 248.50 Matches: 73
Percent Similarity: 48.67% Conservative: 37
Best Local Similarity: 32.30% Mismatches: 90
Query Match: 20.04% Indels: 27
Gaps: 7

US-10-649-273-2_COPY_176_414 (1-239) x US-09-596-002-38 (1-94750)
QY 1 LeuLeuAlaLeuValGInGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20
Db 24305 ATGCTGTGGCGTGGCGATGCGTGTGGCGGTATCATGATTTGGCGAGCTATATCATAT 24364
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40
Db 24365 GCGGTGGGTGAATCTTTGATTAACGCGCAAAATGCTCAAACTG---CCCTATCTCT--- 24418
QY 41 CysSerThrMetSerGlyLysAlaIleGlyHisLeuAlaLysGInGlyAspArgPhe 60
Db 24419 -----GGTGGCCCAAAATTCGAAAAATTAAGCAAAAACGCAACCCACAC 24463
QY 61 HisPheAspIleLysProProLeuHisIleAlaLysAsnGlyAspPheSerPheThrGly 80
Db 24464 GCGTATAGCTGCCAAGACCAATGACAGAT---AAAGGCTGAGTTTTCGTTCTGCTGCGC 24520
```

```
QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGlnLysGlnGlyIleGlyLys 100
Db 24521 ATGAAACCGCATTCATATCATATCAACAGACACCAAAACGCCAAAGCGACCC--- 24577
QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisIleMetAlaCys 120
Db 24578 -----GCCACAGACAGACATCGCCGCAAGCTTGTAGATGCGGTGGTGCAT 24625
QY 121 HisLeuValLysArgThrHisIleArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140
Db 24626 ACTTGTGTCAAAAATATGACCAAAAGCATACAGATACAGGACGCTTGCCAG----- 24676
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAspPheTyIleArgArg 160
Db 24677 -----CTGTGTGTGCGAGGGGGGTCTCTGCGCATGAGATGCTACGCCCG 24721
QY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCySPProProArgLeu 180
Db 24722 ACCCTGACCGAGACGCTCCGCAAAATGATGATGCTGCTGATATATGCTCCCGACG 24781
QY 181 CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGlnArgLeuArgAlaGlyLeu 200
Db 24782 TGCACGATTAATGTGCGATGATCGCTATGCTGCTTTTGTGCTGCTGCTGCTGCTGCA 24841
QY 201 GlyIleLeuHisAspIleGlnGlyIleArgTyArgLupProLysCySPProLeuGlyValAsp 220
Db 24842 TGC-----GATGACTGTGCG-GGTTCGCTGATTC-----CCGATGGATATGAC 24885
QY 221 IleSerLysGlnValGly 226
Db 24886 GACGCTTGCGCTATCGCG 24903

RESULT 7
US-09-540-236-806
; Sequence 806, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATARRHALIS
; FILE REFERENCE: 2709.2005-001
; CURRENT APPLICATION NUMBER: US/09/540,236
; PRIOR FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 806
; LENGTH: 1053
; TYPE: DNA
; ORGANISM: M. catarrhalis
; US-09-540-236-806

Alignment Scores:
Pred. No.: 3,22e-23 Length: 1053
Score: 247.00 Matches: 70
Percent Similarity: 46.85% Conservative: 34
Best Local Similarity: 31.53% Mismatches: 80
Query Match: 19.92% Indels: 38
Gaps: 6

US-10-649-273-2_COPY_176_414 (1-239) x US-09-540-236-806 (1-1053)
QY 1 LeuLeuAlaLeuValGInGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20
Db 445 ATGCTGTGGCGTGGCGATGCGTGTGGCGGTATCATGATTTGGCGAGCTATATCATAT 504
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40
Db 505 GCGGTGGGTGAATCTTTGATTAACGCGCAAAATGCTCAAACTG---CCCTATCTCT--- 558
QY 41 CysSerThrMetSerGlyLysAlaIleGlyHisLeuAlaLysGInGlyAspArgPhe 60
Db 559 -----GGTGTCCCAATATCGAAAAATTAAGCCAAAAACGCAACCCACAC 603
```

```
QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCyAspPheSerPheThrGly 80
Db 604 GCGATGAGCTGCCAGACCGATGACGAT---AAAGGGCTGATTTTTCGTTCACTGAGC 660
QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGlyLysGluGluGlyIleGluLys 100
Db 661 ATGAAAACCGCATTCATTAATCTCATCAAGACAGCCAAAGCCCAAGGACCCCT--- 717
QY 101 GlyIleLeuSerSerAlaAlaAspIleAlaIleAlaThrValGlnHisThrMetAlaCys 120
Db 718 -----GCCACGACGACAGACATCCCGGACGCTTGAGTATGCGGTGATG 765
QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140
Db 766 ACTTGTCGTCAAAATATGACACAGGACTACAGATGACAGGATTCGCGAC----- 816
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheTyrIleArgArg 160
Db 817 -----CTGGTGTGCGACGGGGCGCTCTCGCAATCAGACGCTACGCGCGC 861
QY 161 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLysCysProProArgLeu 180
Db 862 ACCCTGAACCGAGACGCTGCCAAATCGATGCGTGGTGTACTACGCCCGACGAGCTA 921
QY 181 CysThrAspAsnGlyIleMetIleAlaIleArgAsnGlyIleGluArgLeuArgAlaGly--- 199
Db 922 TGACCGGATATATGTTGGGATGATCGCTTACGCTGCTTTTGTCCGTAAGCGGTGGCG 981
QY 200 -----LeuGlyIleLeu 203
Db 982 TCGGATGACTTGGCGGTTCGCTGATATCCCGATGGATATGACGATGCTGTGATCGAA 1041
QY 204 HisAsp 205
Db 1042 TATGAT 1047

RESULT 8
US-09-252-991A-884/C
; Sequence 884, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 884
; LENGTH: 1059
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-884

Alignment Scores:
Pred. No.: 4,81e-21 Length: 1059
Score: 231.00 Matches: 73
Percent Similarity: 49.54% Conservative: 34
Best Local Similarity: 33.80% Mismatches: 87
Query Match: 18.63% Indels: 22
Gaps: 8
DB: 4

US-10-649-273-2_COPY_176_414 (1-239) x US-09-252-991A-884 (1-1059)

QY 2 LeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIleAla 21
Db 621 TTGCTGCGCGGTGAGCGATGCGCGCTACCGACAGTGTGCTTGCGGATCGGTGAGCATGCG 562
QY 22 ProGlyAspMetLeuAspLysValAlaIleArgLeuSerLeuIleLysHisProGluCys 41
```

```
Db 561 GCGCGGAGACCTTGACAGACCGCCAGACTGATCGGCTG---GGCTATCCC----- 511
QY 42 SerThrMetSerGlyGlyValAlaIleGluHisIleLeuAlaLysGlnGlyAsnArgPheHis 61
Db 510 -----GGTGGTCCGGAATTCGCCCGCTGGCGGAGGCGGACCTCTGGCCGC 463
QY 62 PheAspIleLysProProLeuHisHisAlaLysAsnCyAspPheSerPheThrGlyLeu 81
Db 462 TTGCGTTCGCCGCGGCGCATGACCGATCGCCCGGCTGAGCTTCACTTCAAGCGGCTC 403
QY 82 GlnHisValThrAspLysIleIleMetLysLysGlyLysGluGluGlyIleGluLysGly 101
Db 402 AAGACCTTTACCCCTAACC---ACCTGGACGCGCTTGGTGGAGGCGCGCACACGAG 346
QY 102 GlnIleLeuSerSerAlaAlaAspIleAlaIleAlaThrValGlnHisThrMetAlaCysHis 121
Db 345 CAG-----ACCGCTGCGACATCGCTCGGCGCTTCCAGACCGCGGTGTCGAGACC 295
QY 122 LeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuProGln 141
Db 294 CTGCTGATCAATGCGCGTCCGCTTG-----AAGCAGACCGGCGCTG-----AAG 250
QY 142 AsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheTyrIleArgArgAla 161
Db 249 AAC-----CTGCTGATCGCGCGGCTGTACGCGCAACAGCGCTGCGCACGCGC 199
QY 162 LeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLysCysProProArgLeuCys 181
Db 198 CTGGAAGAAGATCTCGCGCAATGAAAGGGCAGGTGTTCTACGCCCGCGGCTTCTGC 139
QY 182 ThrAspAsnGlyIleMetIleAlaIleArgAsnGlyIleGluArgLeuArgAlaGlyLeu 201
Db 138 ACCGACATGCGCGCATGATTCGCTACGCCGCTGCGACGCTGCTCGCGCGC----- 85
QY 202 IleLeuHisAspIleGluGlyIleArgTyrGluProLysCysAspLeu 217
Db 84 ---CAGCATGACGCGCGCGCATGACGCTGCGACGCGCGCTGCGCGCATG 40

RESULT 9
US-09-252-991A-801
; Sequence 801, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 801
; LENGTH: 1206
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-801

Alignment Scores:
Pred. No.: 5,95e-21 Length: 1206
Score: 231.00 Matches: 73
Percent Similarity: 49.54% Conservative: 34
Best Local Similarity: 33.80% Mismatches: 87
Query Match: 18.63% Indels: 22
Gaps: 8
DB: 4

US-10-649-273-2_COPY_176_414 (1-239) x US-09-252-991A-801 (1-1206)

QY 2 LeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIleAla 21
```

Db 604 TTGTCGGCGGTGAGGATATCGCCGCTACCACTTGTGCGCAATCGGCAATCGGCAATCGC 663
Qy 22 ProGlyaspMetLeuaspLysValAlaargargLeuSerLeuIleuLysHisProGluCys 41
Db 664 GCCGCGCAAGCTTGCACAAAGACCGCAAGCTGATGGCTG---GGTATCC----- 714
Qy 42 SerThrmSerGlyGlyValAlaIleGluHisLeuAlaGlnGlyGlnGlyAsnArgPheHis 61
Db 715 -----GGTGTCCGGAATTCGCCCGCTGGCGGACGCGCACTCTGCGCCG 762
Qy 62 PheaspIleLysProPheLeuHisHisAlaLysAsnCyaspPheSerPheThrgLysLeu 81
Db 763 TTCGTCTCCCGCGCGCGATGACCAATCCGCGCGCTGAGCTTCAAGCTTCAAGCGGGCTC 822
Qy 82 GlnHisValThrAspLysIleIleMetLysGlyGlyGluGlyIleGluLysGly 101
Db 823 AAGACCTTTACCTGAC---ACGCGACAGCGTTGGTGTGAGGCGCGGACACACAGCAG 879
Qy 102 GlnIleLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHisThrmAlaCyHis 121
Db 880 CAG-----ACCGCGTCGACATCGCCCTGGCGCTTCAGACCGCGGCTGCGAGACC 930
Qy 122 LeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuProGln 141
Db 931 CTGCTGATCAAGTCCGTCGCGCTTG-----AGCAGACCGCGCTG-----AAG 975
Qy 142 AsnAsnAlaValIleuValAlaSerGlyValAlaSerAspPheThrgLysArgAla 161
Db 976 AAC-----CTGTGATCCCGCGCGGTGTCAGCGCACAGCGCGCTGCGCACGGCC 1026
Qy 162 LeuGluIleLeuThrAsnAlaThrgLysThrgLysLeuCysProProArgLysCys 181
Db 1027 CTGGAAGAAGATGCTCGCGCAATGAAGGCGCAGGTGTTCTACGCGCGCGCGCTTCTC 1086
Qy 182 ThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeu 201
Db 1087 ACCGCAATGGCGCATATCGCTTACGCGCTGCCAGCGCGCTGCTGCCGCGC----- 1140
Qy 202 IleLeuHisAspIleGluGlyIleArgLysTrpLysProLysCysProLeu 217
Db 1141 ---CAGCATGAGCGCGCGCATGACGCTCCAGCGCGCTGCGCGCATG 1185
RESULT 10
US-09-557-884-1/c
; Sequence 1, Application US/09557884
; Patent No. 6506581
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: The Nucleotide sequence of
; the Haemophilus influenzae Rd Genome, Fragments
; thereof, and Uses Thereof
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3 1/2 inch diskette
; COMPUTER: Dell Pentium
; OPERATING SYSTEM: MS DOS v6.22
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/557,884
; FILING DATE: 25-Apr-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 06/476,102
; FILING DATE: JUN-5-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks

; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB186P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-309-8504
; TELEFAX: 301-309-8439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1830121 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1
Alignment Scores:
Pred. No.: 7,8e-15 Length: 1830121
Score: 224.00 Matches: 67
Percent Similarity: 47.50% Conservative: 28
Best Local Similarity: 33.50% Mismatches: 83
Query Match: 18.06% Indels: 22
DB: 4 Gaps: 6
US-10-649-273-2_COPY_176_414 (1-239) x US-09-557-884-1 (1-1830121)
Qy 2 LeuAlaIleuValGlnGlyValSeraspPheLeuLeuGlyLysSerLeuAspIleAla 21
Db 552791 TTAGTCGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 552732
Qy 22 ProGlyaspMetLeuaspLysValAlaargargLeuSerLeuIleuLysHisProGluCys 41
Db 552731 GCTGGCGAAGCTTTATTAACACGCAAAATTACTTGACTA---GATTATCCA----- 552681
Qy 42 SerThrmSerGlyGlyValAlaIleGluHisLeuAlaGlnGlyGlnGlyAsnArgPheHis 61
Db 552680 -----GTCGCGCGGACCTTCTCTGTTAGCGGAAAGTACGCCAAATCGT 552633
Qy 62 PheaspIleLysProPheLeuHisHisAlaLysAsnCyaspPheSerPheThrgLysLeu 81
Db 552632 TTCACATTTCCACGTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 552573
Qy 82 GlnHisValThrAspLysIleIleMetLysLysGlnLysGluGly-----IleGlu 99
Db 552572 AAAACATTTGCCGCAATACAGTTATCAAGCAATTAACAGAGCGGCAACTGATAGAG 552513
Qy 100 LysGlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrgLysGlnHisThrmAla 119
Db 552512 CAA-----ACTAAAGCAGATATTGCTTATGCTTCCAGATGCGGTG 552468
Qy 120 CysHisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeu 139
Db 552467 GATACTCTTGCC-----ATTAAATGTAG---CGTGATTTGAA 552432
Qy 140 ProGlnAsnAsnAlaValIleuValAlaSerGlyValAlaSerAspPheThrgLysLeu 159
Db 552431 GAAACAGCTAATAACGTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 552372
Qy 160 ArgAlaLeuGluIleuThrAsnAlaThrgLysThrgLysLeuCysProProArg 179
Db 552371 GAACGCTTGCCGCACTTAATGCAAAATTAGGTCGGAAGTGTATTCCTCAACCTCA 552312
Qy 180 LeuCysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGly 199
Db 552311 TTTTGTACAGATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 552252
RESULT 11
US-09-643-990A-1/c
; Sequence 1, Application US/09643990A
; Patent No. 6528289
; GENERAL INFORMATION:
; APPLICANT: Robert D. Fleischmann
; Mark D. Adams
; Owen White
; Hamilton O. Smith

J. Craig Venter
TITLE OF INVENTION: The Nucleotide sequence of
the Haemophilus influenzae Rd Genome, Fragments
Thereof, and Uses Thereof
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville,
STATE: MD
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: 3 1/2 inch diskette
COMPUTER: Dell Pentium
OPERATING SYSTEM: MS DOS v6.22
SOFTWARE: ASCII text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/643,990A
FILING DATE: 23-Aug-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/487,429
FILING DATE: 1995-06-07
APPLICATION NUMBER: 08/426,787
FILING DATE: 1995-04-21
ATTORNEY/AGENT INFORMATION:
NAME: Kenley K. Hoover
REGISTRATION NUMBER: 40,302
REFERENCE/DOCKET NUMBER: PB186P1C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-610-5790
TELEFAX: 310-309-8439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1830121 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1
Alignment Scores:
Pred. No.: 7.8e-15 Length: 1830121
Score: 224.00 Matches: 67
Percent Similarity: 47.50% Conservative: 28
Best Local Similarity: 33.50% Mismatches: 83
Query Match: 18.06% Indels: 22
DB: 4 Gaps: 6
US-10-649-273-2_COPY_176_414 (1-239) x US-09-643-990A-1 (1-1830121)
QY 2 LeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIleAla 21
DB 552791 TTACGCGCGTGCAGTGTGTAGAAAATATGAAGATGAGAAATCTTGTGATGTGCT 552732
QY 22 ProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGluCys 41
DB 552731 GCTGGCGAAGCCTTGTGATTAACACGCAAAATTAATTGACTA-----GATTATCCA----- 552681
QY 42 SerThrMetSerGlyGlyValAlaIleGlnHisLeuAlaLysGlnGlyLysAsnArgPheHis 61
DB 552680 -----GGTGGCGCGCGCACCTTCTCGTTTAGCGGAAAAGTACGCCCAAAATCGT 552633
QY 62 PheAspIleLysProPheLeuHisIleAlaLysAsnCysAspPheSerPheThrGlyLeu 81
DB 552632 TTTCACATTTCCACGCTCCATGACAGATCTGCAAGCGCTGATTTAGTTTCTGCTTTA 552573
QY 82 GlnHisValThrAspLysIleIleMetLysGluLysGluGluGly-----IleGlu 99
DB 552572 AAAACATTTGCGCAAAATACAGTATATCAAGCAATTAATAAAGGCGCAACTGATAGAG 552513
QY 100 LysGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAla 119

DB 552512 CAA-----ACTAAGCAGATATTGCTTATGCTTTCCAAAGATGCGGTGG 552468
QY 120 CysHisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeu 139
DB 552467 GATACTCTTGCC-----ATTAATGTAG-----CGTGATTTGAAA 552432
QY 140 ProGlnAsnAsnAlaValLeuValAlaSerGlyValAlaSerAsnPheTyrlleArg 159
DB 552431 GAAACAGCTATAAAGCTTTAGTGTGATGGGAGGGGTAGCGCAATTAATAAATCCCA 552372
QY 160 ArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProPheArg 179
DB 552371 GAAAGCTTGCCGACCTTAATGCMAATTTAGGTGGGAAGTGTATTATCTCAACCTCA 552312
QY 180 LeuCysThrAspAsnGlyLysIleMetIleAlaTPAsnGlyLysGluArgLeuArgIle 199
DB 552311 TTTTGTACAGATTAATGTCGATGATTTGCTTACACAGTATTTTACGTTTAATAACAAGGT 552252
RESULT 12
US-09-902-540-6612
Sequence 6612, Application US/09902540
Patent No. 6833447
GENERAL INFORMATION:
APPLICANT: Goldman, Barry S.
APPLICANT: Goldman, Barry S.
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Wiegand, Roger C.
TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
FILE REFERENCE: 38-10(15849)B
CURRENT APPLICATION NUMBER: US/09/902,540
CURRENT FILING DATE: 2001-07-10
PRIOR APPLICATION NUMBER: 60/217,883
PRIOR FILING DATE: 2000-07-10
NUMBER OF SEQ ID NOS: 16825
SEQ ID NO 6612
LENGTH: 996
TYPE: DNA
ORGANISM: Myxococcus xanthus
US-09-902-540-6612
Alignment Scores:
Pred. No.: 8.47e-20 Length: 996
Score: 221.50 Matches: 70
Percent Similarity: 47.50% Conservative: 25
Best Local Similarity: 35.00% Mismatches: 74
Query Match: 17.86% Indels: 31
DB: 4 Gaps: 8
US-10-649-273-2_COPY_176_414 (1-239) x US-09-902-540-6612 (1-996)
QY 5 ValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIleAlaProGlyAsp 24
DB 433 GTGCAAGCGCTACCGGACGATGACCGCTGTGGGACGACGCGCACACGAGCGCGGAG 492
QY 25 MetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGluCysSerThrMet 44
DB 493 GCATATGACAAAGCCGCTGCATCTCGGCTG-----CCGATCCG----- 534
QY 45 SerGlyGlyLysAlaIleGlnHisLeuAlaLysGlnGlyAsn-----ArgPhe 60
DB 535 ---GGTGGCAGCCCATGACACAGTTGGCCGACAGGGAACCCGAGGCCATCCGCTTC 591
QY 61 HisPheAspIleLysProPheLeuHisIleAlaLysAsnCysAspPheSerPheThrGly 80
DB 592 -----CCGCGCGCGCTGCGCGGCAACAATTCCACGCTGCTTCTCGGG 636
QY 81 LeuGlnHisValThrAspLysIleIleMetLysGluLysGluGluGlyIleGluLys 100
DB 637 TTGAAG-----ACGGCGGTGCTGCACACGTGCAAGGACGGCGTGGCGGAG 684
QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 120

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: November 10, 2005, 16:59:00 ; Search time 793.636 Seconds
(without alignments)
2490.420 Million cell updates/sec

Title: US-10-649-273-2_COPY_176_414
Perfect score: 1240
Sequence: 1 LIALVGVSDFLILGKSLDI.....DISKVGESAIKVPQLKMEI 239

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 9794790 seqs, 4134909567 residues

Total number of hits satisfying chosen parameters: 19589580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=framer.p2n.model -DEV=xip
-Q/cgn2_1/USFTO.spool.p/US10649273/runat_02112005_091339_15609/app.query.fasta_1.1429
-DB=Published Applications_NA -QFMT=fastap -SUFFIX=p2n.rmpb -MINMATCH=0.1
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human0.csl -LIST=45 -DOCCALIGN=200 -THR SCORE=pcr -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10649273 @CGN 1.1 1034 @runat_02112005_091339_15609
-NCPU=6 -ICPU=3 -NO_MMAP -LARGESOURC -NEG_SCORES=0 -WAIT -DSBLCHECK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-Fgapop=6 -Fgapext=7 -Ygapop=10 -Ygapext=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications_NA:*

1: /cgn2_6/prodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/prodata/1/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/prodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/prodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/prodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/prodata/1/pubpna/US07_PUBCOMB.seq.*
7: /cgn2_6/prodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/prodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/prodata/1/pubpna/US09_PUBCOMB.seq.*
10: /cgn2_6/prodata/1/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/prodata/1/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/prodata/1/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/prodata/1/pubpna/US09_PUBCOMB.seq.*
14: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq.*
15: /cgn2_6/prodata/1/pubpna/US10C_PUBCOMB.seq.*
16: /cgn2_6/prodata/1/pubpna/US10C_PUBCOMB.seq.*
17: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq.*
18: /cgn2_6/prodata/1/pubpna/US10C_PUBCOMB.seq.*
19: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq.*
20: /cgn2_6/prodata/1/pubpna/US10C_PUBCOMB.seq.*
21: /cgn2_6/prodata/1/pubpna/US10C_PUBCOMB.seq.*
22: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq.*
23: /cgn2_6/prodata/1/pubpna/US10_NEW_PUB.seq.*
24: /cgn2_6/prodata/1/pubpna/US10_NEW_PUB.seq.*
25: /cgn2_6/prodata/1/pubpna/US10C_PUBCOMB.seq.*
26: /cgn2_6/prodata/1/pubpna/US11_NEW_PUB.seq.*
27: /cgn2_6/prodata/1/pubpna/US60_NEW_PUB.seq.*
28: /cgn2_6/prodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Query Length	DB ID	Description
1	1240	100.0	1416	18	Sequence 177, App
2	1240	100.0	1526	15	Sequence 23, App
3	1240	100.0	1526	19	Sequence 23, App
4	1240	100.0	1526	19	Sequence 23, App
5	1240	100.0	2197	15	Sequence 1, App
6	1240	100.0	2197	19	Sequence 1, App
7	1240	100.0	2197	19	Sequence 1, App
8	1240	100.0	2572	22	Sequence 36, App
9	1213.5	97.9	1387	15	Sequence 21, App
10	1213.5	97.9	1387	19	Sequence 21, App
11	1213.5	97.9	1387	19	Sequence 21, App
12	1213.5	97.9	1387	22	Sequence 21, App
13	1203	97.0	1245	15	Sequence 1047, App
14	1203	97.0	1820	15	Sequence 6, App
15	1059	85.4	2208	18	Sequence 4, App
16	1059	85.4	2890	21	Sequence 400, App
17	725	58.5	14364	15	Sequence 7447, App
18	725	58.5	14364	19	Sequence 20, App
19	725	58.5	14364	19	Sequence 20, App
20	599	48.3	2734	24	Sequence 20, App
21	468	37.7	371	20	Sequence 20426, App
22	468	37.7	371	20	Sequence 3118, App
23	337	27.2	1917	19	Sequence 3118, App
24	337	27.2	1917	19	Sequence 66417, App
25	332.5	26.8	1601	26	Sequence 34189, App
26	270	21.8	1628	20	Sequence 11249, App
27	257	20.7	1146	18	Sequence 14674, App
28	248.5	20.0	1004	20	Sequence 50, App
29	248.5	20.0	1044	18	Sequence 26772, App
30	248.5	20.0	94750	19	Sequence 38, App
31	247	19.9	936	18	Sequence 8315, App
32	241	19.4	1032	18	Sequence 31043, App
33	239	19.3	756	15	Sequence 8, App
34	239	19.3	4360	15	Sequence 2, App
35	234	18.9	300	24	Sequence 7021, App
36	231	18.6	1026	18	Sequence 30016, App
37	231	18.6	1026	18	Sequence 49, App
38	231	18.6	1026	24	Sequence 51, App
39	229	18.5	1026	24	Sequence 6946, App
40	224	18.1	1029	9	Sequence 22020, App
41	224	18.1	1830121	18	Sequence 1, App
42	224	18.1	1830121	21	Sequence 1, App
43	224	18.1	1830121	24	Sequence 1, App
44	224	18.1	1830121	21	Sequence 529, App
45	217	17.5	9967	24	Sequence 529, App

ALIGNMENTS

RESULT 1
US-10-120-988-177
; Sequence 177, Application US/10120988
; Publication No. US20030219745A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Goodrich, Ryle
; APPLICANT: Liu, Chenghua
; APPLICANT: Ren, Feiyun
; APPLICANT: Wang, Dunrui
; APPLICANT: Dmanac, Radoje T.
; TITLE OF INVENTION: No. US20030219745A1 Nucleic Acids and
; FILE REFERENCE: 802CON

```

; CURRENT APPLICATION NUMBER: US/10/120,988
; CURRENT FILING DATE: 2002-04-11
; PRIOR APPLICATION NUMBER: 09/774,528
; PRIOR FILING DATE: 2001-01-30
; NUMBER OF SEQ ID NOS: 441
; SOFTWARE: pc_FL_genes Version 2.0
; SEQ ID NO 177
; LENGTH: 1416
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (205) ..(1305)
; US-10-120-988-177

Alignment Scores:
Pred. No.: 2,38e-151 Length: 1416
Score: 1240.00 Matches: 239
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 18 Gaps: 0

US-10-649-273-2_COPY_176_414 (1-239) x US-10-120-988-177 (1-1416)
QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlySerLeuAsp1le 20
Db 586 CTGTTGGCATTACTTCAAGAGATTTCAGATTTCTGCTTGGAAAGCTTTGGACATA 645
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeu1lelyshisProGlu 40
Db 646 GCACCAAGTGACATGCTTGAACAAGTGGCAAGACCTTTCTTAATAAATGATCCAGAG 705
QY 41 CysSerThrMetSerGlyGlyValAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60
Db 706 TGGTCCACCAATGAGTGGTGGAAAGCCATAGAACATTGGCCAAACAAAGAAATGATTT 765
QY 61 HisPheAsp1leLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80
Db 766 CATTGGACATCAAACTCCCTTGATCATGCTAAATTTGTGATTTTCTTTTACTGGA 825
QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGlyIleGluLys 100
Db 826 CTTCAACACGTTACTGATAAATATATATGAAAAAGAAAGAAAGATTTGAGAG 885
QY 101 GlyGlnIleLeuSerSerAlaAlaAsp1leAlaIleAlaThrValGlnHisThrMetAlaCys 120
Db 886 GGGCAAAATCTGCTTTCAGACAGACAGATGCTGCCACAGTACACACAAATGCAATGT 945
QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuPro 140
Db 946 CATCTGTGAAAAGAACATCGGGCTATTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1005
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheTyrIleArgArg 160
Db 1006 CAAATATATGCAATGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTG 1065
QY 161 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeu 180
Db 1066 GCTCTGGAATTTTAAACAAACGCAACAGTGCATTTGTGTGTGTGTGTGTGTGTGTGT 1125
QY 181 CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeu 200
Db 1126 TGCATGATATATGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1185
QY 201 GlyIleLeuHisAsp1leGluGlyIleArgTyrGluProLysCysProLeuGlyValAsp 220
Db 1186 GGCATTTTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1245
QY 221 HisSerLysGluValGlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 239
Db 1246 ATATCAAAAAGAGTGGAGAGCTTCATATAAGTACACAAATTAATAATGAGATA 1302
```

```

RESULT 2
US-10-067-443-23
; Sequence 23, Application US/10067443
; Publication No. US20030082782A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 23
; LENGTH: 1526
; TYPE: DNA
; ORGANISM: homo sapiens
; US-10-067-443-23

Alignment Scores:
Pred. No.: 2.65e-151 Length: 1526
Score: 1240.00 Matches: 239
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 15 Gaps: 0

US-10-649-273-2_COPY_176_414 (1-239) x US-10-067-443-23 (1-1526)
QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlySerLeuAsp1le 20
Db 85 CTGTTGGCATTACTTCAAGAGATTTCAGATTTCTGCTTGGAAAGCTTTGGACATA 144
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeu1lelyshisProGlu 40
Db 145 GCACCAAGTGACATGCTTGAACAAGTGGCAAGACCTTTCTTAATAAATGATCCAGAG 204
QY 41 CysSerThrMetSerGlyGlyValAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60
Db 205 TGGTCCACCAATGAGTGGTGGAAAGCCATAGAACATTGGCCAAACAAAGAAATGATTT 264
QY 61 HisPheAsp1leLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80
Db 265 CATTGGACATCAAACTCCCTTGATCATGCTAAATTTGTGATTTTCTTTTACTGGA 324
QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGlyIleGluLys 100
Db 325 CTTCAACACGTTACTGATAAATATATGAAAAAGAAAGAAAGATTTGAGAG 384
QY 101 GlyGlnIleLeuSerSerAlaAlaAsp1leAlaIleAlaThrValGlnHisThrMetAlaCys 120
Db 385 GGGCAAAATCTGCTTTCAGACAGACAGATGCTGCCACAGTACACACAAATGCAATGT 444
QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuPro 140
Db 445 CATCTGTGAAAAGAACATCGGGCTATTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 504
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheTyrIleArgArg 160
Db 505 CAAATATATGCAATGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGAT 564
QY 161 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeu 180
Db 565 GCTCTGGAATTTTAAACAAACGCAACAGTGCATTTGTGTGTGTGTGTGTGTGTGTGT 624
QY 181 CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeu 200
Db 625 TGCATGATATATGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 684
```



```
QY      201 GYIleuHISaSPllleGluGlyIleArgTYrGluProlyeCyseProleuGlyValAsp 220
DB      685 GGCATTTTACATGACATAGAGGATCCGCTATGAAACCAAAATGCTCTTGGAAGTAAAC 744
QY      221 IleserlysgluValGlyGluAlaSerileysValProGlnleuLyseMetGluile 239
DB      745 ATATCAAAAAGAAAGTTGAGAAAGCTTCCATTAAGTACCACATTTAAAAATGGAGATA 801

RESULT 3
US-10-649-273-23
; Sequence 23, Application US/10649273
; Publication No. US20040043407A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1
; FILE REFERENCE: D0073 CNT
; CURRENT APPLICATION NUMBER: US/10/649, 273
; PRIOR FILING DATE: 2003-08-27
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 23
; LENGTH: 1526
; TYPE: DNA
; ORGANISM: homo sapiens
US-10-649-273-23

Alignment Scores:
Pred. No.:      2,65e-151      Length:      1526
Score:          1240.00      Matches:      239
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    100.00%      Indels:      0
DB:             19          Gaps:          0

US-10-649-273-2_COPY_176_414 (1-239) x US-10-649-273-23 (1-1526)
QY      1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLyseSerleuAspIle 20
DB      85 CTGTGGCATTAAGTTCAAGAGGATTTCTGCTTCTTGAAAGTCTTTGGACATA 144
QY      21 AlaProGlyAspMetLeuAspLyseValAlaArgArgLeuSerleuIleLyseHisProGlu 40
DB      145 GCACCAAGGTGACATGCTTGCAAGGTGGCAAGAACTTTCTTAATAAATCATCCAGAG 204
QY      41 CysSerThrMetSerGlyGlyLyseAlaIleGluHisleuAlaLyseGlnGlyAsnArgPhe 60
DB      205 TGTCTCCACCAAGAGTGTGGGAAAGCCATGAAACATTTGGCAAAACAAAGAAATAGATTT 264
QY      61 HisPheAspIleLyseProProleuHisHisAlaLyseAsnCyAspPheSerPheThrGly 80
DB      265 CATTTCGACATCAAACTCCCTTCGATCATGCTAAAAATTTGATTTTCTTTTACTGGA 324
QY      81 LeuGlnHisValThrAspLyseIleIleMetLyseLyseGluLyseGluGluGlyIleGluLyse 100
DB      325 CTTCAACACGTTACTGATTAATAATATGAAAAAGAAAAAGAGAAAGTATTGAGAAAG 384
QY      101 GLYGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 120
DB      385 GGGCAAAATCTGCTTCACAGACAGACATTCGCCACAGTACGCAACACATGGCAGTGT 444
QY      121 HisLeuValLyseArgThrHisArgAlaIleLeuPheCysLyseGlnArgAspLeuLeuPro 140
DB      445 CATCTTGTAAGAAAGACACATCGGGCTATTCTGTTTGTAAGCAAGAGACTTGTACTCT 504
QY      141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheTYrIleArgArg 160
DB      505 CAAATAATATGCACTACTGTTGATCTGTGTGTCTGCCAAGTAACTTTATATATCCGACGA 564
```

```
QY      161 AlaLeuGluIleLeuThrAsnAlaThrGlnCyseThrLeuCyseProProArgLeu 180
DB      565 GCTCTGCAAAATTTTAAACAAACGACACAGTGCATCTTGTGTGCTCTCCACAGACTA 624
QY      181 CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeu 200
DB      625 TGCATCTGATTAAGGATTAATGATTCATGGAATGGATTTGAAAGAACTAGTGTGCTGGCTTG 684
QY      201 GYIleuHISaSPllleGluGlyIleArgTYrGluProlyeCyseProleuGlyValAsp 220
DB      685 GGCATTTTACATGACATAGAGGATCCGCTATGAAACCAAAATGCTCTTGGAAGTAAAC 744
QY      221 IleserlysgluValGlyGluAlaSerileysValProGlnleuLyseMetGluile 239
DB      745 ATATCAAAAAGAAAGTTGAGAAAGCTTCCATTAAGTACCACATTTAAAAATGGAGATA 801

RESULT 4
US-10-651-722-23
; Sequence 23, Application US/10651722
; Publication No. US20040048302A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1
; FILE REFERENCE: D0073 DIV
; CURRENT APPLICATION NUMBER: US/10/651,722
; PRIOR FILING DATE: 2003-08-29
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 23
; LENGTH: 1526
; TYPE: DNA
; ORGANISM: homo sapiens
US-10-651-722-23

Alignment Scores:
Pred. No.:      2,65e-151      Length:      1526
Score:          1240.00      Matches:      239
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    100.00%      Indels:      0
DB:             19          Gaps:          0

US-10-649-273-2_COPY_176_414 (1-239) x US-10-651-722-23 (1-1526)
QY      1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLyseSerleuAspIle 20
DB      85 CTGTGGCATTAAGTTCAAGAGGATTTCTGCTTCTTGAAAGTCTTTGGACATA 144
QY      21 AlaProGlyAspMetLeuAspLyseValAlaArgArgLeuSerleuIleLyseHisProGlu 40
DB      145 GCACCAAGGTGACATGCTTGCAAGGTGGCAAGAACTTTCTTAATAAATCATCCAGAG 204
QY      41 CysSerThrMetSerGlyGlyLyseAlaIleGluHisleuAlaLyseGlnGlyAsnArgPhe 60
DB      205 TGTCTCCACCAAGAGTGTGGGAAAGCCATGAAACATTTGGCAAAACAAAGAAATAGATTT 264
QY      61 HisPheAspIleLyseProProleuHisHisAlaLyseAsnCyAspPheSerPheThrGly 80
DB      265 CATTTCGACATCAAACTCCCTTCGATCATGCTAAAAATTTGATTTTCTTTTACTGGA 324
QY      81 LeuGlnHisValThrAspLyseIleIleMetLyseLyseGluLyseGluGluGlyIleGluLyse 100
DB      325 CTTCAACACGTTACTGATTAATAATATGAAAAAGAAAAAGAGAAAGTATTGAGAAAG 384
QY      101 GLYGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 120
```

```
Db 385 GGGCAAAATCTGCTTCAGCAGACACATTCCTCCACAGTACACACACAAATGGCATGT 444
Qy 121 HisleuValIysArgThrHisArgAlaIleleuPheCysIysGlnArgAspLeuPro 140
Db 445 CATCTTGAAAAAGAACACATCGGGCTATTCTGTTTGTAAAGCAGAGACTTGTACTT 504
Qy 141 GlnAsnAsnAlaValIeuValAlaSerGlyGlyValAlaSerAsnPheTyrIleArgArg 160
Db 505 CAATAATATGCAGTACTGCTGTGACATCGTGCTGCGCAAGTAACTTATATCCGACGA 564
Qy 161 AlaIeuGluIleleuThrAsnAlaThrGlnCysThrIleuLeuCysProProProArgLeu 180
Db 565 GCCTCGTGAATTTTAAACAAACGCAACAGTGCACCTTGTGTCTCTCCACAGACTA 624
Qy 181 CysThrAspAsnGlyIleMetIleAlaThrAsnGlyIleGluArgLeuArgAlaGlyLeu 200
Db 625 TGCACGTATATGGCATTTATGATTCATGATGATGATGATGATGATGATGATGATGATG 684
Qy 201 GlyIleleuHisAspIleGluGlyIleArgTyrGluProIlyCysProIeuGlyValAsp 220
Db 685 GGCATTTTACATGACATATGAAAGCATCGCTATGACCAAAATGTCCCTTGGAGTAAAC 744
Qy 221 IleserIysGluValIleGlyIleValIleSerIleIysValProGlnIleuIysMetGluIle 239
Db 745 ATATCAAAAAGAGTTGAGAAAGCTTCATATAAAGTACCACAATTAAATAATGAGATA 801

RESULT 5
US-10-067-443-1
; Sequence 1, Application US/10067443
; Publication No. US20030082782A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE HIGHLY EXPRESSED I
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 2197
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (231)..(1472)
US-10-067-443-1

Alignment Scores:
Pred. No.: 4,52e-151 Length: 2197
Score: 1240.00 Matches: 239
Percent Similarity: 100.00% Conservativeness: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0
DB: 15

US-10-649-273-2_COPY_176_414 (1-239) x US-10-067-443-1 (1-2197)

Qy 1 LeuValAlaValAlaGlnGlyValIleSerAspPheLeuLeuGlyIleSerIleuAspIle 20
Db 756 CTGTGGCATTAAGTTCAGAGAGTTCAGATTTTCGCTTCTTGGAAAGCTCTTGGACATA 815
Qy 21 AlaProGlyAspMetIleuAspIysValAlaArgArgLeuSerIleIleIysHisProGlu 40
Db 816 GCACACAGGTGACATGCTTGAACAGGTGCAAGAACACTTCTTAAATAAACATCCAGAG 875
Qy 41 CysSerThrMetSerGlyIleValAlaIleGluHisIleuAlaIysGlnGlyAsnArgPhe 60
Db 876 TGCTCCACACAGATGATGGAGAAAGCCATAGAACATTTGGCCAAACAAAGAAATGATTT 935
```

```
Qy 61 HisPheAspIleIysProProIleuHisHisAlaIysAsnCysAspPheSerPheThrGly 80
Db 936 CATTTGACATCAAACTCCCTTGATCATGTCTAAATAATGGATTTTCTTTTACTGGA 995
Qy 81 LeuGlnHisValThrAspIysIleIleMetIysIysGluIysGluGlyIleGluIys 100
Db 996 CTTCAACACGTTACTGATTAATATATATGAAAAAGAAAAAGAGAAAGATTTTGAAG 1055
Qy 101 GlyGlnIleIeuSerSerAlaIleAspIleAlaIleAlaThrValGlnHisThrMetAlaCys 120
Db 1056 GGGCAAAATCTGCTTCAGCAGACACATTCGTCACAGTACAGACACACAAATGGCATGT 1115
Qy 121 HisleuValIysArgThrHisArgAlaIleleuPheCysIysGlnArgAspLeuPro 140
Db 1116 CATCTTGAAAAAGAACACATCGGGCTATTCTGTTTGTAAACAGAGACTTGTACTT 1175
Qy 141 GlnAsnAsnAlaValIeuValAlaSerGlyGlyValAlaSerAsnPheTyrIleArgArg 160
Db 1176 CAATAATATGACGATCTGCTGTGACATCGTGCTGCGCAAGTAACTTATATCCGACGA 1235
Qy 161 AlaIeuGluIleleuThrAsnAlaThrGlnCysThrIleuLeuCysProProProArgLeu 180
Db 1236 GCTCGGAAATTTTAAACAAACGCAACAGTGCACCTTGTGTCTCTCCACAGACTA 1295
Qy 181 CysThrAspAsnGlyIleMetIleAlaThrAsnGlyIleGluArgLeuArgAlaGlyLeu 200
Db 1296 TGCACGTATATGGCATTTATGATTCATGGAATGATGATGATGATGATGATGATGATG 1355
Qy 201 GlyIleleuHisAspIleGluGlyIleArgTyrGluProIlyCysProIeuGlyValAsp 220
Db 1356 GGCATTTTACATGACATATGAAAGCATCGCTATGACCAAAATGTCCCTTGGAGTAAAC 1415
Qy 221 IleserIysGluValIleGlyIleValIleSerIleIysValProGlnIleuIysMetGluIle 239
Db 1416 ATATCAAAAAGAGTTGAGAAAGCTTCATATAAAGTACCACAATTAAATAATGAGATA 1472

RESULT 6
US-10-649-273-1
; Sequence 1, Application US/10649273
; Publication No. US20040043407A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1
; FILE REFERENCE: D0073 CNT
; CURRENT APPLICATION NUMBER: US/10/649,273
; PRIOR FILING DATE: 2003-08-27
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 2197
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (231)..(1472)
US-10-649-273-1

Alignment Scores:
Pred. No.: 4,52e-151 Length: 2197
Score: 1240.00 Matches: 239
Percent Similarity: 100.00% Conservativeness: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0
DB: 19

US-10-649-273-2_COPY_176_414 (1-239) x US-10-649-273-1 (1-2197)
```

Oy		1	LeuleuAlaleuValGInGIyVAlSeRASPheLeuLeuGIyLysSerLeuAspIle	20
Db		756	CtGTGGcATTAGTTCaAGAGcTTTCAGATTTTCGCTTCtTGAAAGcTTTGACATA	815
Oy		21	AlaProGIYASpMeTLeuAspLysValAlaArgAqLeuSerLeuIlleySHIsProGU	40
Db		816	GCACCAAGGTGAcATCCTCGACAAGGTGGCCAAAGAACCCTTCTTAATAAAAATCATCCAGAG	875
Oy		41	CysSerTHmetSeSGlyGLyValAlaleGIWHIsleuAlalyGIInGIyAsnArgHe	60
Db		876	TGCTCACCATAGAGTGGTGGGAACCCATAGAACAATTGGCCAAACAGAGAAATGAAATT	935
Oy		61	HISpheAspILeYProProLeuHISHIsAlalyAsnCysAspPheSerPhethrIGLy	80
Db		936	CATTITGACATCAACACTCCCTCGCATCATGCTAAAAATTTGATTTTCTTTTCTGCA	995
Oy		81	LeuGINHisValThrasPLysILElleMetLysGLyUGLyGLyILleGIUlys	100
Db		996	CTTCAACACGTTACTGATATAATATAAGAAAAGAAAAGAAAGAAAGATTAGAAG	1055S
Oy		101	GlyGINIlleLeuSerSerILalAlasplllealaatrvalGINHIsThmeChALcyS	120
Db		1056	GGGCAAAATCCGTGCTTCAGACAGCAACATTCGTGCCAGATGACACACAATGGCCATGT	1115S
Oy		121	HISleuValLYsArgThHIsARgAlalleuPheCYSLysGINArgAspLeuEuro	140
Db		1116	CATCTTGGAAGAAAGAACATCGGGCTATTCTGTTTTGTAGCAAGAGACTGTGTACT	1175S
Oy		141	GINsnAsnaAlavalLeuValAlaserGIyValAlaserAspHetyrILLEArgArG	160
Db		1176	CAAAATAATGcAGTAcTGCTGGTGCATCTGTGGTGGCGAAGTAATTATATCCGACGA	1235S
Oy		161	AlAleuGIUlleLeuthAsnAlathrgINCysThrLeuLeuCysProProArgLeu	180
Db		1236	GCTCTGGAATTTTAAACAAACGCAACAAGTGCACCTTGTGTGTCCTCCCTCCAGACTA	1295S
Oy		181	CysThrAspANgIyleMeTlIleAlatrpAnngIyleGIuArgLeuArgAlAGLyEu	200
Db		1286	TGCATGTATTAATGGCATATATATATGcATGAGATGGATATTAAGAAAGCTACGTGGCTTG	1355S
Oy		201	GLyIlleLeuHIsApILeGIULyILEArgTYrGIUProLysCysProLeuGIyValAsp	220
Db		1356	GGCATTTTACATGACATGAGGAAGGcATCCGCTATGAACCAAATATGCTCTTGAGAGTAgAC	1415S
Oy		221	IleSerLYsgIUvalGIyGluAlaserIlleyValProGINLeuLYsmetGIUlle	239
Db		1416	ATATCAAAAGAGTGGAGAGACTTCCATAAAAGTACACCAATTAAAAATGAGAGTA	1472
RESULT 7				
US-10-651-722-1				
:	Sequence 1, Application US/10651722			
:	Publication No. US20040048302A1			
:	GENERAL INFORMATION:			
:	APPLICANT: Bristol-Myers Squibb Company			
:	TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METRALOPROTEASE, MP-1			
:	FILE REFERENCE: D00073 DIV			
:	CURRENT APPLICATION NUMBER: US/10/651,722			
:	CURRENT FILING DATE: 2003-08-29			
:	PRIOR APPLICATION NUMBER: US 60/266,518			
:	PRIOR FILING DATE: 2001-02-05			
:	PRIOR APPLICATION NUMBER: US 10/067,443			
:	PRIOR FILING DATE: 2002-02-05			
:	PRIOR APPLICATION NUMBER: US 60/282,814			
:	PRIOR FILING DATE: 2001-04-10			
:	NUMBER OF SEQ ID NOS: 71			
:	SOFTWARE: PatentIn version 3.2			
:	SEQ ID NO 1			
:	LENGTH: 2197			
:	TYPE: DNA			
:	ORGANISM: Homo sapiens			
:	FEATURE:			
:	NAME/KEY: CDS			

```

US-10-651-722-1
/ LOCATION: (231) . (1472)
US-10-651-722-1

Alignment Scores:
Pred. No.: 4,52e-151 Length: 2197
Score: 1240.00 Matches: 239
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 19 Gaps: 0

US-10-649-273-2_COPY_176_414 (1-239) x US-10-651-722-1 (1-2197)

QY 1 LeuLeuAlaLeuAlaValSerAspPheLeuLeuGlyValSerLeuAsp1le 20
DB 756 CTTGTTGGCATTTAGTTCAGAGAGTTTCAATTTTCGCTTTGGAGAAAGCTTTGGACATA 815
QY 21 AlaProGlyAspMetLeuAspValAlaArgArgLeuSerLeuIleValHisProGlu 40
DB 816 GCACGAGGTGACATGCTTGACAGAGTGGCAAGAGACTTTCTTTATATAAACATTCAGAG 875
QY 41 CysSerThrMetSerGlyValValAlaIleGluHisLeuAlaValGlnGlyAsnArgPhe 60
DB 876 TGCTTCACCATGAGAGGTGGGAAAGCCATAGAACATTGGCCCAACAGAAATGATTT 935
QY 61 HisPheAspIleValProProLeuHisHisAlaIysAsnGlyAspPheSerPheThrGly 80
DB 936 CATTTTGCATCAAACTCCCTCCATCATGCTATAAAATTTGATTTTCTTTTAACTGGA 995
QY 81 LeuGlnHisValIleThrAspValIleIleMetLysLysGlyValGlnGlyValIleGluLys 100
DB 996 CTTGACACGCTTACGTATGATTAATATATAGAAAAAGAAAGAGAGATTTGAGAG 1055
QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleAlaThrValGlnHisThrMetAlaCys 120
DB 1056 GGGCAAAATCTGTCTTGACAGACAGACTTGTCTGCCACATGACAGACAAATGGCANTGT 1155
QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuPro 140
DB 1116 CATCTTGGAAAAACACATCGGGCTAATTCGTTTGTATGACAGAGACTGTCTTACT 1175
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyValAlaSerAsnPheThrIleArgArg 160
DB 1176 CAATATATGACAGTACGTGTGTCATCGTGCGTGGCAATATACCTTATATCCGAGGA 1235
QY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180
DB 1236 GCTCGGAAATTTTAAACAAGCAACACAGTGGCTTTGTGTCTCTCCCAAGACTA 1295
QY 181 CysThrAspAsnGlyIleMetIleAlaIleTrpAsnGlyIleGluArgLeuArgAlaGlyLeu 200
DB 1296 TGCACTGATATATGCAATTATGATGTCATGAGAAATGATTTGAAAGACTAACGTCGGCTTG 1355
QY 201 GlyIleLeuHisAspIleGluGlyIleAlaArgValProLysCysProLeuGlyValAsp 220
DB 1356 GGCATTTTACATGACATAGAGAGCATCGCATATACCAAAATATCTCTTGGAGTGAAC 1415
QY 221 IleSerLysGluValAlaGlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 239
DB 1416 ATATCAAAAGAGTGGAGAGACTTCCATATAAGTACCAATTAATAATGAGATA 1472

RESULT 8
US-10-480-388-36
/ Sequence 36, Application US/10480988
/ Publication No. US2005069877A1
/ GENERAL INFORMATION:
/ APPLICANT: GANDHI, Ameena R.; KABLE, Amy E.;
/ APPLICANT: SWARNAKAR, Anita; HAFALIA, April J.A.;
/ APPLICANT: TRAN, Bao; DUGGAN, Brendan M.;
/ APPLICANT: WARREN, Bridget A.; ISON, Craig H.;
/ APPLICANT: HONGHELL, Cynthia D.; NGUYEN, Daniel B.;
/ APPLICANT: LU, Dzung Alma M.; LEE, Ernestine A.;
/ APPLICANT: YUE, Henry; FORSYTHE, Ian J.;

```

```
APPLICANT: BARROSO, Ines; RAMKUMAR, Jayalaxmi;
APPLICANT: GRIFFIN, Jennifer A.; Li, Joana X.;
APPLICANT: YANG, Junming; THANAGAVELU, Kavitha;
APPLICANT: GIETZEN, Kimberly J.; DING, Li;
APPLICANT: BAUGHN, Marian R.; BOROMSKY, Mark L.;
APPLICANT: YAO, Monica G.; CHAMLA, Nandinder K.;
APPLICANT: MASON, Patricia M.; GURURAJAN, Rajagopal;
APPLICANT: LEE, Sally; BECHTA, Shanya D.;
APPLICANT: LEE, Soo Yeun; TRAN, Uyen K.;
APPLICANT: ELLIOTT, Vicki S.; LUO, Wen;
APPLICANT: SPRAGUE, William W.; TANG, Y. Tom;
APPLICANT: LU, Yan; ZEBARJADIAN, Yeganeh
TITLE OF INVENTION: PROTEIN MODIFICATION AND MAINTENANCE MOLECULES
FILE REFERENCE: PF-1040 USN
CURRENT APPLICATION NUMBER: US/10/480,988
CURRENT FILING DATE: 2003-12-16
PRIOR APPLICATION NUMBER: PCT/US02/19360
PRIOR FILING DATE: 2002-06-18
PRIOR APPLICATION NUMBER: US 60/300,508
PRIOR FILING DATE: 2001-06-22
PRIOR APPLICATION NUMBER: US 60/303,445
PRIOR FILING DATE: 2001-07-06
PRIOR APPLICATION NUMBER: US 60/305,405
PRIOR FILING DATE: 2001-07-13
PRIOR APPLICATION NUMBER: US 60/311,442
PRIOR FILING DATE: 2001-08-09
PRIOR APPLICATION NUMBER: US 60/314,821
PRIOR FILING DATE: 2001-08-24
PRIOR APPLICATION NUMBER: US 60/315,992
PRIOR FILING DATE: 2001-08-29
PRIOR APPLICATION NUMBER: US 60/378,205
PRIOR FILING DATE: 2002-05-03
NUMBER OF SEQ ID NOS: 56
SOFTWARE: PERL Program
SEQ ID NO 36
LENGTH: 2572
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
OTHER INFORMATION: Incyte ID No: 7632424CB1
US-10-480-988-36

Alignment Scores:
Pred. No.: 5.7e-151 Length: 2572
Score: 1240.00 Matches: 239
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 22 Gaps: 0

US-10-649-273-2_COPY_176_414 (1-239) x US-10-480-988-36 (1-2572)
QY 1 LeuLeuAlaLeuValGInGInValISerAspPheLeuLeuGInGInYLySSerLeuAspIle 20
DB 669 CTGTTGGCACTTGAAGGAGTTTCAGATTTCCTCTTGGAAAGTTTGGACATA 728
QY 21 AlaProGlyAspMetLeuAspIlyValAlaArgArgLeuSerLeuIleYshISProGlu 40
DB 729 GCACCAAGTGCATGTTGACAAAGTGCAGAAAGCTTTCTTAATAAACAATCCAGAG 788
QY 41 CysSerThrMetSerGlyGlyValAlaIleGInHISLeuAlaYsGInGInYAsnArgPhe 60
DB 789 TGCTCCACCATGAGTGTGGGAAAGCCATAGAACTTTGGCCAAACAGAAATAGATT 848
QY 61 HisPheAspIleYsProProLeuHISHisAlaYAsnCyAspPheSerPheThrGly 80
DB 849 CATTGTGACATCAACCTCCCTTGCATCATGCTAAATAATGTATTTCTTTACTGGA 908
QY 81 LeuGInHISValThrAspIlySileIleMetYsLYsGInGInYLySInGInGInYIleGInLYs 100
DB 909 CTTCAACACCTTACTATTAATATATATATCAAAAGAAAGAAAGAAAGTATTTGAGAG 966
```

```
QY 101 GlyGInHISLeuSerSerAlaAlaAspIleAlaIleThrValGInHISThrMetAlaCys 120
DB 969 GGGCAAAATCTGCTTACAGACAGCATTTGGTCCACAGTACAGACCAATGGCATGT 1028
QY 121 HisLeuValIleArgThrHisArgAlaIleLeuPheCysLYsGInArgAspLeuPro 140
DB 1029 CATCTTGTAAAGAAACACATCGGCTATTCGTTTGTGAAGAGACTGTGTACCT 1088
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheYrIleArgArg 160
DB 1089 CAATAATATGCACTACTGTTGATTCGTTGCTGGCAAGTAACTTATATCCGAGA 1148
QY 161 AlaLeuGInHISLeuThrAsnAlaThrGInCysThrLeuLeuCysProProArgLeu 180
DB 1149 GCTCTGGAATTTTAAACAAGCAACAGTGCATTTGTTGTGCTCTCCACAGACTA 1208
QY 181 CysThrAspAsnGlyIleMetIleAlaTPAsnGlyIleGInArgLeuAlaGlyLeu 200
DB 1209 TGCACCTGATATGTCATTTGATTCATGGAATGTAATGAAGACTACGTGCTG 1268
QY 201 GlyIleLeuHISAspIleGInGInYIleArgYrGlyProLYsCysProLeuGlyValAsp 220
DB 1269 GGCATTTTACATACATAGAAAGCATCCGCTATGAACCAAAATGCTCTTGGAGTAC 1328
QY 221 IleSerLYsGInValGlyGInAlaSerIleYsValProGInLeuYsMetGInIle 239
DB 1329 ATATCAAAAGAACTTGGAGAGCTTCATTAAGTACCAATATTAATGAGAGATA 1385

RESULT 9
US-10-067-443-21
; Sequence 21, Application US/10067443
; Publication No. US20030082782A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 21
; LENGTH: 1387
; TYPE: DNA
; ORGANISM: homo sapiens
US-10-067-443-21

Alignment Scores:
Pred. No.: 6.82e-148 Length: 1387
Score: 1213.50 Matches: 238
Percent Similarity: 90.15% Conservative: 0
Best Local Similarity: 90.15% Mismatches: 1
Query Match: 97.86% Indels: 25
DB: 15 Gaps: 1

US-10-649-273-2_COPY_176_414 (1-239) x US-10-067-443-21 (1-1387)
QY 1 LeuLeuAlaLeuValGInGInValISerAspPheLeuLeuGInGInYLySSerLeuAspIle 20
DB 549 CTGTTGGCACTTGAAGGAGTTTCAGATTTCCTCTTGGAAAGTCTTTGGACATA 608
QY 21 AlaProGlyAspMetLeuAspIlyValAlaArgArgLeuSerLeuIleYshISProGlu 40
DB 609 GCACCAAGTGCATGCTTGAACAAGTGCAGAAAGACTTTCTTAATAAACAATCCAGAG 668
QY 41 CysSerThrMetSerGlyGlyValAlaIleGInHISLeuAlaYsGInGInYAsnArgPhe 60
DB 669 TGCTCCACCATGAGTGTGGGAAAGCCATAGACATTTGGCCAAACAGAAATAGATT 728
```


PRIOR APPLICATION NUMBER: US 10/067,443
PRIOR FILING DATE: 2002-02-05
PRIOR APPLICATION NUMBER: US 60/282,814
PRIOR FILING DATE: 2001-04-10
NUMBER OF SEQ ID NOS: 71
SOFTWARE: PatentIn version 3.2
SEQ ID NO 21
LENGTH: 1387
TYPE: DNA
ORGANISM: homo sapiens
US-10-651-722-21

Alignment Scores:

Score: No.:	6,82e-148	Length: 1387
Percent Similarity:	1213.50	Matches: 238
Best Local Similarity:	90.15%	Conservative: 0
Query Match:	97.86%	Mismatches: 1
		Indels: 25
		Gaps: 1

US-10-649-273-2_COPY_176_414 (1-239) x US-10-651-722-21 (1-1387)

QY 1 LeuEua1aleuValgInglyValSerAspPheLeuLeuGlyLySerLeuAsp1le 20
DB 549 CTGTTGCATTAGTTCAAGAGAGTTTCAGATTTCGCTTCTTGAAAGCTTTGGACATA 608
QY 21 AlaProGlyAspMetLeuAspLyValAlaArgArgLeuSerLeu1lelyshisProGlu 40
DB 609 GCACAGGTGACATGCTTGACAGAGTGCAAGAACATCTTTTAATAAATCCAGAG 668
QY 41 CysSerThrMetSerGlyGlyValAlaIleGluHisLeuAlaLyGlnGlyAsnArgPhe 60
DB 669 TGCTCCACCATGAGTGTGGGAAAGCCATAGACATTTGGCCAAACAGAAATGATT 728
QY 61 HisPheAsp1leLySProProLeuHisIshAlaLyAsnCyAspPheSerPheThrgly 80
DB 729 CATTTGACATCAAACTCCCTTGATCATGCTAAATTTGATTTTCTTTACTGGA 788
QY 81 LeuGlnHisValThrAspLyS1leIleMetLySlySgluLySgluGly1le----- 98
DB 789 CTTCAACACGTTACTGATTAATAATATGAAAAAGAAAAAGAGATATATTCTTA 848
QY 98 ----- 98
DB 849 ATTAGTAAGTTGAACAGATTAATATTCCTGATTTGCTCTAAATTAAGCTCATTTTC 908
QY 99 -----GluLySgluGlnIleLeuSerSerAlaAlaAsp1leAlaIleThrValGln 115
DB 909 TGCAGGTATGAGAGGGGCAAAATCCTGTCTTCAGCAGCAGACATTTGCTCCACAGTACAG 968
QY 116 HisThrMetAlaCyShiSleuValIlySArgThrHisArgAlaIleLeuPheCySlySglu 135
DB 969 CACACAATGGCATGCTCTTGTGAAAAAGAACACATCGGGCTATTCTGTTTGTAAAGCAG 1028
QY 136 ArgAspLeuLeuProGlnAsnAsnAlaValIleuValAlaSerGlyGlyValAlaSerAsn 155
DB 1029 AGAAGCTTTGTAACCTCAAAATAATATGACAGTGTGCTGATGTGTGTGCGAAGTATAC 1088
QY 156 PheTyrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCySThrLeuLeuCyS 175
DB 1089 TTCTATATCCGAGAGCTCTGGAATTTTAAACAAACGACACAGAGCATTTGTGTGTGT 1148
QY 176 ProProProArgLeuCySThrAspAsnGlyIleMetIleAlaIleAsnGlyIleGluArg 195
DB 1149 CCTCCTCCAGACATATGACATGATATGATGATATGATGATGATGATGATGATGATG 1208
QY 196 LeuArgAlaGlyLeuGlyIleLeuHisAsp1leGluGlyIleArgTyrGluProLyCyS 215
DB 1209 CTACGT 1268
QY 216 ProLeuGlyValAsp1leSerLySgluValIleGlyValAlaSerIleLySValProGlnLeu 235
DB 1269 CCTCTTGAGTAGATATCAAAAGAGAGTGTGAGAGCTTCCATTAAGTACCAATTTA 1328

QY 236 LysMetGluIle 239
DB 1329 AAAATGAGAGATA 1340

RESULT 12

US-10-887-553A-1047
Sequence 1047, Application US/10887553A
Publication No. US20050085436A1
GENERAL INFORMATION:

APPLICANT: Garza, Dan
APPLICANT: Li, Hao
TITLE OF INVENTION: Method to treat conditions associated
with insulin signalling dysregulation
FILE REFERENCE: 4-33262
CURRENT APPLICATION NUMBER: US/10/887,553A
CURRENT FILING DATE: 2004-07-08
PRIOR APPLICATION NUMBER: 60/485,883
PRIOR FILING DATE: 2003-08-07
NUMBER OF SEQ ID NOS: 1208
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1047
LENGTH: 1387
TYPE: DNA
ORGANISM: human
US-10-887-553A-1047

Alignment Scores:

Pred. No.:	6,82e-148	Length: 1387
Score:	1213.50 <td>Matches: 238</td>	Matches: 238
Percent Similarity:	90.15% <td>Conservative: 0</td>	Conservative: 0
Best Local Similarity:	97.86% <td>Mismatches: 1</td>	Mismatches: 1
Query Match:		Indels: 25
		Gaps: 1

US-10-649-273-2_COPY_176_414 (1-239) x US-10-887-553A-1047 (1-1387)

QY 1 LeuEua1aleuValgInglyValSerAspPheLeuLeuGlyLySerLeuAsp1le 20
DB 549 CTGTTGCATTAGTTCAAGAGAGTTTCAGATTTCGCTTCTTGAAAGCTTTGGACATA 608
QY 21 AlaProGlyAspMetLeuAspLyValAlaArgArgLeuSerLeu1lelyshisProGlu 40
DB 609 GCACAGGTGACATGCTTGACAGAGTGCAAGAACATCTTTTAATAAATCCAGAG 668
QY 41 CysSerThrMetSerGlyGlyValAlaIleGluHisLeuAlaLySgluGlyIle----- 98
DB 669 TGCTCCACCATGAGTGTGGGAAAGCCATAGACATTTGCTCAAAACAGAAATGATT 728
QY 61 HisPheAsp1leLySProProLeuHisIshAlaLyAsnCyAspPheSerPheThrgly 80
DB 729 CATTTGACATCAAACTCCCTTGATCATGCTAAATTTGATTTTCTTTACTGGA 788
QY 81 LeuGlnHisValThrAspLyS1leIleMetLySlySgluLySgluGly1le----- 98
DB 789 CTTCAACACGTTACTGATTAATAATATGAAAAAGAAAAAGAGATATATTCTTA 848
QY 98 ----- 98
DB 849 ATTAGTAAGTTGAACAGATTAATATTCCTGATTTGCTCTAAATTAAGCTCATTTTC 908
QY 99 -----GluLySgluGlnIleLeuSerSerAlaAlaAsp1leAlaIleThrValGln 115
DB 909 TGCAGGTATGAGAGGGGCAAAATCCTGTCTTCAGCAGCAGACATTTGCTCCACAGTACAG 968
QY 116 HisThrMetAlaCyShiSleuValIlySArgThrHisArgAlaIleLeuPheCySlySglu 135
DB 969 CACACAATGGCATGCTCTTGTGAAAAAGAACACATCGGGCTATTCTGTTTGTAAAGCAG 1028
QY 136 ArgAspLeuLeuProGlnAsnAsnAlaValIleuValAlaSerGlyGlyValAlaSerAsn 155
DB 1029 AGAAGCTTTGTAACCTCAAAATAATATGACAGTGTGCTGATGTGTGTGCGAAGTATAC 1088

```
Qy 156 PheTyrIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThrIleuLeuCys 175
Db 1089 TTTCTATATCCGACAGCTCTGGAAATTTTAAACCAACGACAGTGCATTGTGTGT 1148
Qy 176 ProProAlaArgLeuCysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArg 195
Db 1149 CTTCTCCGACGACTATGACATGATATGATTCATGATGGAATGGTATTGAACA 1208
Qy 196 LeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrGluProLysCys 215
Db 1209 CTACGTGTGGCTTGGGCAATTTTACATGACATAGAACATCCGCTATGAACCAAAATGT 1268
Qy 216 ProLeuGlyValaAspIleSerIleGluValaGlyIleValaSerIleValaProGluLeu 235
Db 1269 CCTCTTGAGTAGACATATCAAAAGAAAGTGGAGAGCTTCCATAAAGTACCAAAATTA 1328
Qy 236 LysMetGluIle 239
Db 1329 AAAATGAGATA 1340

RESULT 13
US-10-012-140-6
; Sequence 6, Application US/10012140
; Publication No. US20030009017A1
; GENERAL INFORMATION:
; APPLICANT: Leiby, Kevin R.
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Glucksmann, Maria A.
; TITLE OF INVENTION: 38650, 28472, 5495, 65507, 81588, AND
; TITLE OF INVENTION: 14354 METHODS AND COMPOSITIONS OF HUMAN PROTEINS AND USES
; FILE REFERENCE: THEREOF
; CURRENT APPLICATION NUMBER: US/10/012,140
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: 60/246,768
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/246,772
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/249,185
; PRIOR FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 1245
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-012-140-6

Alignment Scores:
Pred. No.: 1,38e-146 Length: 1245
Score: 1203.00 Matches: 232
Percent Similarity: 98.33% Conservative: 3
Best Local Similarity: 97.07% Mismatches: 4
Query Match: 97.02% Indels: 0
DB: 15 Gaps: 0

US-10-649-273-2_COPY_176_414 (1-239) x US-10-012-140-6 (1-1245)
Qy 1 LeuLeuAlaLeuValaGlnGlyValaSerAspPheLeuLeuLeuGlyLysSerIleuAspIle 20
Db 526 CTGTGGCATATAGTTCAAGAGAGTTTCAGATTTTCGCTTCTTGAAAGTCTTTGGACATA 585
Qy 21 AlaProGlyAspMetLeuAspLysValaAlaArgArgLeuSerIleuIleLysHisProGlu 40
Db 586 GCACCAAGTAGACATGCTTGACAGAGTGCAGAAAGACTTTCTTAATTAACATCCAGAG 645
Qy 41 CysSerThrMetSerGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60
Db 646 TGCCTCCACCATGAGTGGTGGAAAGCCATAGAACCTTTGGCCAAACAAAGAAATGATTT 705
Qy HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80
Db 706 CATTTCACATCAAACTCCCTTCATCATCTAAATAATTGTGATTTTCTTTACTGGA 765
```

```
Qy 81 LeuGlnHisValaThrAspLysIleIleMetLysGluLysGluGluGlyIleGluLys 100
Db 766 CTTCAACACGTTACTGATTAATAATATGAAGAAACAGAAACAGAGAGAGATTTGAGAAG 825
Qy 101 GlyGlnIleLeuSerSerIleAlaAspIleAlaIleValaGlnHisThrMetAlaCys 120
Db 826 GGGCAAAATCTGTCTTCACACACACATTCGTCACAGTACAGACACAAATGGCATGT 885
Qy 121 HisLeuValaLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140
Db 886 CATCTTGCAAAAGAAACATATCGGCTATTCTTTTGTAGAGAGACATTTGTACT 945
Qy 141 GlnAsnAsnAlaValaLeuValaLysSerGlyIleValaAlaSerAsnPheTyrIleArgArg 160
Db 946 CAAATTAATGACGACTGCTGTCATCTGATGTGGTGGCAAGTAATCTTATATCCGACAG 1005
Qy 161 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProAlaGlu 180
Db 1006 GCTCTGAAATTTTAAACCAACGACACAGTGCATCTTGTGTCTCTCCACACTA 1065
Qy 181 CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlu 200
Db 1066 TGCACCTGATATGCACTTATGATTCATGATGAAATGGTATTGAAGACTATGCTGCTTG 1125
Qy 201 GlyIleLeuHisAspIleGluGlyIleArgTyrGluProLysCysProLeuGlyValaAsp 220
Db 1126 GCATTTTACATGACATAGAAAGCATCCGCTATGAACCAAAATGCTCTTGGAGTAGAC 1185
Qy 221 IleSerLysGluValaGlyGluLysSerIleLysValaProGlnLeuLysMetGluIle 239
Db 1186 ATATCAAAAGAAAGTGGAGAGAGCTTCATTAAGTACCAATTAATAATGAGATA 1242

RESULT 14
US-10-012-140-4
; Sequence 4, Application US/10012140
; Publication No. US20030009017A1
; GENERAL INFORMATION:
; APPLICANT: Leiby, Kevin R.
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Glucksmann, Maria A.
; TITLE OF INVENTION: 38650, 28472, 5495, 65507, 81588, AND
; TITLE OF INVENTION: 14354 METHODS AND COMPOSITIONS OF HUMAN PROTEINS AND USES
; FILE REFERENCE: THEREOF
; CURRENT APPLICATION NUMBER: US/10/012,140
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: 60/246,768
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/246,772
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/249,185
; PRIOR FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1820
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (146)...(1390)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1820)
; OTHER INFORMATION: n = A,T,C or G
US-10-012-140-4

Alignment Scores:
Pred. No.: 2,41e-146 Length: 1820
Score: 1203.00 Matches: 232
Percent Similarity: 98.33% Conservative: 3
Best Local Similarity: 97.07% Mismatches: 4
```

Query Match: 97.02% Index: 0
DB: 15 Gaps: 0
US-10-649-273-2_copy_176_414 (1-239) x US-10-012-140-4 (1-1820)

```
QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyValSerLeuAsp11e 20
DB 671 CTGTTGGCATTAATGCAAGAGTTTCAGATTTCGTCTTCTTGAAAGTCTTTGGACATA 730
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40
DB 731 GACCCAGGTACATGCTTGACAGAGTGCAGAGAACCTTCTTTAAATAACATCCAGAG 790
QY 41 CysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60
DB 791 TGCCTCCACCATGAGTGTGGGAAAGCCATAGACATTTGGCCAAACAGAAATAGATT 850
QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCyAspPheSerPheThrGly 80
DB 851 CATTTTGACATCAAACTCCCTTGCAATGCTAAATAATGATTTTCTTTTACTGGA 910
QY 81 LeuGlnHisValThrAspLysIleIleMetLysGlyGlyGluGluGlyIleGluLys 100
DB 911 CTTCACACGTTACTGATTAATAATGAAAAACAGAAACAGAGAGATTGAGAG 970
QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 120
DB 971 GGGCAAAATCCTGCTTTCAGCAGCAGACATTCCTCCACAGTACAGACACATGCGCAT 1030
QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140
DB 1031 CATCTTGAAAAAGAACACATCGGGCTATTCTGTTTGTAGCAGAGACTTGTACT 1090
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaAsnAsnPheTyrlleArgArg 160
DB 1091 CAAATATATCAGTACGTGGTTCATCTGGTGTCCGAACTTCTATATCCGACGA 1150
QY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180
DB 1151 GCTCTGAAATTTTAACAAACGCAACACAGTGCATTTGTGTCTCTCCACGACTA 1210
QY 181 CysThrAspAsnGlyIleMetIleAlaIleArgAsnGlyIleGluArgLeuArgAlaGlyLeu 200
DB 1211 TGCACGTATATGCAATTAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATG 1270
QY 201 GlyIleLeuHisAspIleGluGlyIleArgGlyGluProLysCysProLeuGlyValAsp 220
DB 1271 GGCATTTTACATGACATAGAGAGCATCGCTATGAAACCAAAATGCTCTTGAGATGAC 1330
QY 221 IleSerTyrsGluValGlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 239
DB 1331 ATATCAAAAGAGTTGGAGAGCTTCAATAAAGTACCACAATTAATAATGAGATA 1387

RESULT 15
US-10-094-749-400
; Sequence 400. Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KETICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
```

```
; APPLICANT: OTSUKA, MOTOKYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 08435/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 400
; LENGTH: 2208
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-094-749-400
```

Alignment Scores:
Pred. No.: 2,3e-127 Length: 2208
Score: 1059.00 Matches: 211
Percent Similarity: 88.28% Conservative: 0
Best Local Similarity: 88.28% Mismatches: 4
Query Match: 85.40% Indels: 24
DB: 18 Gaps: 1

US-10-649-273-2_copy_176_414 (1-239) x US-10-094-749-400 (1-2208)

```
QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyValSerLeuAsp11e 20
DB 869 CTGTTGGCATTAATGCAAGAGTTTCAGATTTCGTCTTCTTGAAAGTCTTTGGACATA 928
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40
DB 929 GACCCAGGTACATGCTTGACAGAGTGCAGAGAACCTTCTTTAAATAACATCCAGAG 988
QY 41 CysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60
DB 989 TGCCTCCACCATGAGTGTGGGAAAGCCATAGAACATTTGGCCAAACAGAAATAGATT 1048
QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCyAspPheSerPheThrGly 80
DB 1049 CATTTTGACATCAAACTCCCTTGCAATGCTAAATAATGATTTTCTTTTACTGGA 1108
QY 81 LeuGlnHisValThrAspLysIleIleMetLysGlyGlyGluGluGlyIleGluLys 100
DB 1109 CTTCACACGTTACTGATTAATAATGAAAAAGAGAGAGATTTGAGAG 1168
QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 120
DB 1169 GGGCAAAATCCTGCTTTCAGCAGCAGACATTCCTCCACAGTACAGACACATGCGCAT 1228
QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140
DB 1229 CATCTTGAAAAAGAACACATCGGGCTATTCTGTTTGTAGCAGAGACTGTTTACT 1288
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaAsnAsnPheTyrlleArgArg 160
DB 1289 CAAATATATGCAGTCTGTTTCATCTGGTGTGCAGAGTACTTCTGTATCCGACGA 1348
QY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180
DB 1349 GCTCTGAAATTTTAACAAACGCAACACAGTGCATTTGTGTCTCTCCACACATA 1408
QY 181 CysThrAspAsnGlyIleMetIleAlaIleArgAsnGlyIleGluArgLeuArgAlaGlyLeu 200
DB 1409 TGCACGTATATGCAATTAATGATTTGCA----- 1435
QY 201 GlyIleLeuHisAspIleGluGlyIleArgGlyGluProLysCysProLeuGlyValAsp 220
DB 1436 -----TGATGCTCTTGAGATGAGAC 1456
```


OY	221	IleSerLysGluValaGlGluAlaSerIleLysValProGlnLeuLysMetCgluIle	239
DQ	1457	ATATCAAAAGAGTGGAGAAGCTCCATTAAAAGTACCAACTTAATAATGCAGATA	1513

Search completed: November 11, 2005, 03:16:43
Job time : 803.636 secs